AD-A082 491 RAND CORP SANTA MONICA CA F/G 6/5
QUALITY OF CARE PROVIDED BY PHYSICIAN'S EXTENDERS IN AIR FORCE --ETC(U)
JAN 80 G A GOLDBERG, D G JOLLY F49620-77-C-0023
UNCLASSIFIED RAND/R-2436-AF NL

END
DATE FILMED
5-80
DTIC

March 10, 1980

PUBLICATIONS DEPARTMENT

ERRATUM

R-2436-AF Quality of Care Provided by Physician's Assistants in Air Force Primary Medicine Clinics, George A. Goldberg and David G. Jolly

In Table 1 on page 8, it should have been noted that the first line "Total number of MDs," reflects the number of physicians who saw patients during the four-week sample taken at each demonstration base. This number includes the supervising team physicians (three at Chanute, four at Dyess, three at Fairchild, and four at Nellis). It also includes primary care physicians not included in the team structure and other specialists who occasionally assume duties in the emergency room or walk-in clinics.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER R-2436-AF	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Quality Of Care Provided By Physician's Extenders In Air Force Primary Medicine Clinics		5. TYPE OF REPORT & PERIOD COVERED Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G. A. Goldberg, D. G. Jolly		8. CONTRACT OR GRANT NUMBER(s) F49620-77-C-0023
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Rand Corporation 1700 Main Street Santa Monica, California 90401		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Requirements, Programs & Studies Group (Af/RDQM), DCS/R&D and Acquisition Hq USAF, Wash., D. C. 20330		12. REPORT DATE January 1980 13. NUMBER OF PAGES 76
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

No Restrictions

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Air Force
Medical Services
Nursing
Physicians

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Evaluates the quality of care of physician's extenders (23 physician's assistants and 7 primary care nurse practitioners) in Air Force primary medicine clinics, as part of an evaluation of PEs assuming care formerly provided by physicians. Physician's assistants performed at least as well as MDs on 25 out of 28 nonredundant process-of-care criteria. Nurse practitioners met the MD standard on 14 of 19 criteria. No major differences were found in PEs' use of ancillary services (laboratory and x-ray) or orders for further care when controlling for case-mix. As expected, PEs consulted MDs infrequently, but more often for serious complaints and at rates similar to those found in other PE studies. The study concludes that the Air Force can deliver the same quality of care when PEs treat a sizeable proportion of patients formerly treated by MDs. 78 pp. (Author).

RAND/R-2436-AF
January 1980

Quality of Care Provided by Physician's Extenders in Air Force Primary Medicine Clinics

George A. Goldberg, David G. Jolly

A Project AIR FORCE report
prepared for the
United States Air Force

F49620-77-C-0012

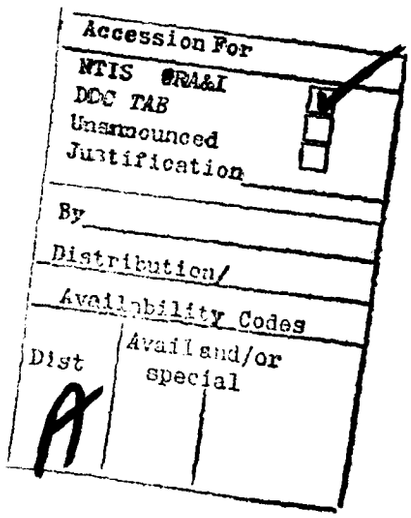
Rand
SANTA MONICA, CA. 90406

PREFACE

In 1976-1978, as part of a study of "Air Force Health Delivery Systems" for Project AIR FORCE, Rand assisted the Air Force Surgeon General with a demonstration project in the use of physician's extenders in primary medicine outpatient clinics. This report, one of a series presenting Rand's evaluation of the demonstration project, assesses the quality of care delivered by the extenders. Two other reports are forthcoming: the first will provide an overview of the project, focusing on operational and economic issues; the second will describe patients' attitudes toward physician's extenders and the demonstration project.

In addition to its immediate audience in the office of the Surgeon General, the present report should be of interest to the wider medical community, since it documents the quality of care that can be achieved with physician's extenders in a favorable institutional environment.

Other topics addressed by the health-delivery-systems project include the supply of physicians and the demand for outpatient services. The research is conducted for Project AIR FORCE under Rand's Manpower, Personnel, and Training Program.



SUMMARY

With the end of the physician draft, the Air Force has experienced a decrease in the number of available physicians, particularly for outpatient (ambulatory) care. In 1976, Rand proposed alleviating that shortage by introducing large numbers of physician's extenders (physician's assistants and primary care nurse practitioners) into the Air Force setting; these extenders might assume a sizable proportion of the patient load previously treated by physicians. Agreeing with the potential merit of Rand's recommendations, the Air Force established a demonstration project in 1976 at four base hospitals to test the feasibility of using physician's extenders (PEs) to provide a large portion of primary care in the clinics. This report presents an evaluation of the quality of care provided during the demonstration project, by comparing PEs' performance with that of physicians working in the same setting.

We adduced several strands of evidence in evaluating the medical appropriateness of having PEs assume a substantial portion of the workload. We assembled the results of condition-specific quality-of-care criteria applied to the care provided; reviewed differences in the pattern of ordering tests, procedures, and return visits; and analyzed the supervisory-consultative relationships between physicians and extenders.

We find that the Air Force can deliver the same quality of care when PEs treat a sizable percentage of patients formerly treated by physicians. We first compared PEs' and physicians' treatment of common, simple, outpatient problems for five classes of quality-of-care criteria defining desirable diagnostic actions, desirable therapeutic actions, undesirable diagnostic actions, undesirable therapeutic actions, and desirable disposition (follow-up) actions. We found that physician's assistants (PAs) performed as well as or better than physicians in 25 out of 28 criteria, and primary care nurse practitioners (PCNPs) performed as well as or better than physicians in 14 out of 19 criteria. For most criteria, we found no significant differences among the three groups of providers, nor could we identify any systematic deficiency on the part of any group of providers.

In addition, there was no overall difference in the performance of PAs and PCNPs. Furthermore, we did not find a decay in PAs' performance when 1974 and 1977 data were compared, nor did we identify differences between recently graduated and earlier graduated PAs. Thus, the "product" being graduated from the Air Force's own, in-house training program appears to be both consistent and stable.

We found that PEs ordered diagnostic procedures at rates similar to those of physicians. Some differences in ordering rates were observed, but there was no consistent evidence for significant overburdening of the Air Force's care delivery system. Analysis of return visits, referrals, and hospital admissions provided further evidence of appropriate extender behavior.

Finally, supervision of the extenders by physicians occurred with reasonable frequency, and was more likely to occur for cases judged as complex or serious. This suggests that extenders' care was adequately supervised. Consultation took up only a modest amount of the physicians' time.

In all these aspects—quality, utilization, consultation—PEs perform satisfactorily. Insofar as we can determine, extenders measure up to the performance level of physicians working in the same setting. Their performance in the demonstration project also compares well with previous studies of extender quality, even in light of the higher ratio of PEs to physicians at the demonstration bases.

We believe that the Air Force can continue to employ PEs in outpatient clinics without diminution in the quality of care. Indeed, these findings medically justify this expanded use of PEs in primary care settings throughout the Air Force.

ACKNOWLEDGMENTS

We acknowledge with gratitude the intellectual leadership and support of David S. C. Chu, who directed this research project from its inception in 1973 until 1978.

We wish to thank Susan Hosek and C. Robert Roll for their continuous and valuable contributions to our work. Able programming assistance was provided by Leola Cutler, Kathy Scofield, and Andrew Siegel. William Lisowski also helped with a number of statistical issues. We especially thank our reviewers, Sheldon Greenfield and Emmett Keeler, for their constructive suggestions.

We are indebted to numerous Air Force personnel for their cooperation and support. The Office of the Surgeon General established the demonstration project and generously assisted us in the evaluation. Fred Ippoliti, our project monitor, has been especially helpful. The staffs of the demonstration hospitals at Chanute, Dyess, Fairchild, and Nellis Air Force Bases implemented the project, collected the data analyzed in this report, and spent a considerable amount of time answering our questions. In particular, we note with appreciation the contributions of the four hospital commanders: Col. Gilbert Kitching, Col. Blair Behringer, Col. Thomas Coolidge, and the late Col. William Walter.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS vii

Section

I. INTRODUCTION 1
- Background 1
- The Demonstration Project 1
- Physician's Extenders 2
- Outline of the Study 3

II. DATA AND METHODS 4
- Quality of Care Measurement 4
- Relative Quality of Care 4
- The Patient Contact Record (Encounter Form) as a Data Collection Instrument 5
- Selection of Condition-Specific Criteria 6
- Other Measurements of Quality 7
- Sample Considerations 8

III. CONDITION-SPECIFIC CRITERIA 9
- Background 9
- Selection of Criteria 9
- Classes of Criteria 9
- Representative Criteria 11
- Breadth of Criteria 11
- Summary Comparison 12
- Comparison of 1977 and 1974 Results 15
- Conclusions 16

IV. UTILIZATION ANALYSIS 18
- Introduction 18
- Methods 18
- Ordering for all Conditions 18
- Ordering for Specific Conditions 19
- Ordering of X-rays 20
- Ordering of Physical Therappy 22
- Ordering of Electrocardiograms 22
- Ordering of Urinalyses and Other Tests 22
- Rates of Ordering Tests: Application 22
- Disposition in General 25
- Conclusions 28

V. SUPERVISION OF PHYSICIAN'S EXTENDERS 29

VI. CONCLUSION 34

Appendix

A. PATIENT CONTACT RECORD AND CONDITION FREQUENCIES 37
B. STATISTICAL TECHNIQUES 48
C. DISCUSSION OF SELECTED CRITERIA OF QUALITY 59
D. RELATIONSHIP BETWEEN UTILIZATION AND QUALITY: LITERATURE REVIEW 71
E. DIAGNOSTIC CATEGORIES 73

BIBLIOGRAPHY 75

I. INTRODUCTION

BACKGROUND

Since the end of the physician draft, fewer physicians have been available to the Air Force. The shortage has been, and will continue to be, particularly acute in outpatient (ambulatory) care, a traditional arena for general medical officers (GMOs) who are in especially short supply.

In 1976, Rand proposed alleviating the physician shortage by introducing large numbers of physician's extenders (PEs) into the Air Force setting, where they might assume a sizable portion of the patient load previously treated by physicians. Although the PE concept might have merit under any circumstances, it appeared especially important because of the shortage.

Rand recommended that the outpatient primary medicine areas be staffed by teams of providers consisting of one physician and two or three highly-trained PEs (the term includes both physician's assistants (PAs) and Primary Care Nurse Practitioners (PCNPs)).[1] In this way, the Air Force might deal with the shortage of primary medicine physicians while providing high-quality and responsive medical care in a cost-effective manner. Because the recommendations required extrapolation beyond actual experience with PEs in the Air Force, and because the Air Force did not have extenders to implement the concept on a large scale, Rand recommended a demonstration program at a small number of installations. The Surgeon General approved this recommendation.

In the fall of 1976, the Air Force began the demonstration project in the primary medicine clinics of four base hospitals (Chanute, Dyess, Fairchild, and Nellis) to evaluate a care-delivering system using a large number of PEs. The purpose of this report (one in a series concerning the demonstration project) is to examine the quality of care provided by PEs during the demonstration.

The report assesses the quality of primary medical care when a substantial portion of it is delivered by PCNPs and PAs, in concert with physicians. For this assessment, we use a comparative method; that is, we compare the PEs' quality of care with that of physicians, at least for the type of routine outpatient problems that PAs and PCNPs are trained to treat.

We present the results of specific quality-of-care criteria applied to the care that PEs provide; review the differences in the way they order tests, procedures, and return visits; and analyze the supervisory-consultative relationships between physicians and PEs. Several strands of evidence are thus available for the comparison.

THE DEMONSTRATION PROJECT

As organizational steps, the demonstration project proceeded to:

- Enrich the mix of PEs in primary medicine clinics, principally using PAs;
- Organize providers into teams, with a physician as supervisor;
- Use PEs to provide first echelon of care. Physician serves as a supervisor, consultant, referral point, and regular provider for previously seen complex patients.
- Organize patients (including retiree families) into panels, with one team treating one panel;

[1]The term "physician's extenders" is controversial, and will be discussed below together with a description of PAs and PCNPs.

- Establish a system where all patients except true emergencies are seen by appointment;
- Furnish (rationed) direct access to a physician for those (few) patients who strongly prefer a physician;
- Retain other usual system features (personnel rotation, facilities, support personnel).

The most important project feature was the high ratio of PEs to primary care physicians: either two or three to one. This ratio was based on a 1974 study, conducted at nine Air Force installations. Because of their greater availability, the project included more PAs than PCNPs.

The providers were organized into teams that replaced the traditional primary care clinics (general therapy and flight surgeons).[2] In these teams, the physician played a supervisory role. Most outpatients initially saw an extender. The physician was responsible for monitoring extenders' care, furnishing consultation, caring directly for the more difficult patients, and providing general medicine inpatient care.

With the teams organized in this fashion, it was possible to do something about one of the persistent complaints concerning military medicine: its impersonal nature, typical of many institutional systems. Patients were assigned in panels so that each active and retired household using the base for care had an identified team of practitioners.

Although the project contained important elements of the family practice approach, it was not identical to it, and in fact offered some additional flexibility. Team chiefs were not limited to family practitioners, although family practitioners and internists were favored. The demonstration bases also retained the traditional Obstetrics/Gynecology and Pediatric clinics. The division of labor between the teams and the Obstetrics/Gynecology and Pediatric clinics depended on their relative richness of staffing. In practice, because of shortages of pediatrics personnel, the teams cared for all except young children.

PHYSICIAN'S EXTENDERS

The term "physician's extenders" is controversial. Other authors prefer to use the term "new health professionals," or "allied health professionals," but virtually all workers in this field, including us, agree that no single term yet devised is entirely appropriate. PEs are non-physicians who perform some of the medical and administrative tasks traditionally performed by physicians. In its broadest sense, the term "physician's extender" conceivably could refer to receptionists (who have often performed triage over the telephone) in physicians' offices, medical corpsmen or technicians, and nurse anesthetists, among others. This report, however, uses the term, "physician's extenders" to refer to only two types of new health professionals: physician's assistants and nurse practitioners.[3]

The Air Force PA is a former corpsman who has graduated from a two-year program of instruction operated by the Air Force. The program includes one year of classroom work in the basic sciences, and a one-year rotation through the outpatient departments of a large Air Force hospital. The PA is therefore trained to diagnose and treat common illnesses, and can also help manage complex patient problems under the supervision of a physician.

The Air Force nurse practitioner (NP) is a registered nurse who has taken a variable amount of additional training, similar to that received by the PA, but usually less extensive.

---

[2]Depending on the physician's training, they could also absorb the workload of internal medicine.

[3]AMOSISTS (volunteer corpsmen who receive two weeks' classroom instruction followed by ten weeks of on-the-job training, and who then deliver care by strict adherence to printed algorithms as part of the so-called Automated Military Outpatient System) are also excluded from our definition of the term physician's extenders, even though they perform in military medical settings (Vickery et al., 1975).

Like the PA, the NP is trained to diagnose and treat common illnesses and to help manage complex patient problems under a physician's supervision. Unlike the PAs in the Air Force, however, who have functioned mainly as generalists, NPs have often been specially trained in obstetrics/gynecology and pediatrics, as well as primary care.

No branch of the military uses PAs and PCNPs in precisely the same way, but in the Air Force they are virtually interchangeable. Despite the differences in both their background and training, they are expected to handle similar types of patients with similar types of diseases in similar settings. It is reasonable, then, to consider them jointly in our analysis, although we also examine their performance separately.

Numerous studies examining selected aspects of performance in different settings have shown that PEs can provide a portion of the care formerly provided by physicians, without a sacrifice in the level of quality. The reader is referred to the sweeping evaluation by Appel and Lowin (1975), as well as to the more specific reports of Perry (1977), Lewis et al. (1969), Sackett et al. (1974), and Levine et al. (1976). If one wishes to look specifically at the technical process of care rendered by PEs, limited data are available in studies by Komaroff (1974), Tompkins et al. (1977), Sibley et al. (1975), and Duttera and Harlan (1978). There has been one published study (Kane, Olsen, and Castle, 1976) in which PEs' technical process of care was compared with that provided by their physician preceptors, through the use of an encounter form (described below) similar to that used in our work.

Our study reexamines the issue of the quality of care that can be provided by PEs. It focuses on the technical process of care in the military outpatient medical setting. The demonstration project's high ratio of extenders to physicians raises the question of maintenance of quality under this condition. The great absolute number of extenders, differences in their training, and variations in the amount of experience they have had (possibly affecting the quality of their current practice)—these are other important reasons for measuring quality of care within the context of the demonstration project.

OUTLINE OF THE STUDY

Section II of this report presents the methods of data collection and measurement. Section III explores five categories of condition-specific criteria that measure quality, and analyzes the results. Section IV presents a utilization analysis of ordering rates for tests and procedures. It also offers a brief analysis of the clinic's self-generation of visits. Section V examines supervision of PEs, a measure of system quality. We examine the consultation and disposition patterns for physicians and PEs. Section VI presents our general conclusions.

II. DATA AND METHODS

QUALITY OF CARE MEASUREMENT

Quality of care criteria are judgmental statements regarding what is to be considered "good and bad" medical care. Medical care can be viewed as consisting of "technical" aspects (what is done or not done to the patient, and how it is done) and "caring" aspects (humaneness of interaction, establishment of rapport between the patient and the provider, etc.) (Howell, 1976; Goldberg, 1977). The data source, the Patient Contact Record (described below), dictates that our quality of care criteria deal solely with the technical aspects of care. We are also generally limited to evaluation of the "process" of care rather than the "outcome" of care (Donabedian, 1966). Process measures concentrate on the procedures, methods, and strategies of care, while outcome measures focus on the effects of treatment (Kane, 1976). These topics have been thoroughly reviewed (Brook, 1973; Brook and Williams, 1976; Williams and Brook, 1978; Greenfield et al., 1977), and will not be discussed here.

The general limitations on evaluating the technical quality of care offer few problems in the general outpatient setting, where conditions are relatively straightforward (though not necessarily well-defined). In the outpatient setting, many presenting conditions treated are common and conventionally managed, and there is a relatively clear connection between the process of care and its outcome (cure or symptom relief). Therefore, we can use such criteria to evaluate the care rendered by physician's extenders (PEs) and to understand how that care compares and contrasts with the care rendered by physicians working in similar settings.

Quality of care criteria were selected to cover a range of age groups; acute and chronic diseases; preventive, diagnostic, and therapeutic services; and physical and mental concerns.

RELATIVE QUALITY OF CARE

Quality of care criteria need not be applied as all-or-nothing (100 percent or 0 percent) compliance standards. "Absolute quality" values of 100 percent or 0 percent (compliance with a criterion) are unlikely to represent the best or worst possible care, because patients always vary enough to render absolute values questionable. For example, a provider who decides to check the box "coryza" (runny nose) as the diagnosis, even though the patient also complains of a mild sore throat, may still decide to swab the throat to obtain a throat culture. The record will then show that throat cultures were obtained from some patients with "coryza", while our criterion may specify that no "coryza" patients need throat cultures, but patients with "pharyngitis" (sore throat) do. Where does that leave our criterion? We believe that the criterion remains intact—its relativity a virtue, not a flaw. Given two groups of patients with the diagnoses just mentioned, for example, it is a reasonable criterion to say that a substantially higher percentage of pharyngitis patients should have throat cultures taken.

How high (or how low) should percentages of compliance be? In view of such problems as multiple (or unrecorded) diagnoses, variations in disease severity, and other factors, arbitrary numbers cannot be justified. However, so long as we can state that "higher" is "better," we can still make meaningful statements about the *relative* quality of care, which is what we want to do. We have the further advantage of wishing to compare *groups* of providers: one physician group with one extender group, or one extender group with another. Therefore, we compare

their performances according to their compliance rates for each criterion.[1]

The gist of the comparative method of evaluating quality of care is therefore: two (or more) groups to compare; a quality of care criterion; and a compliance standard which, while not absolute, does specify if better quality of care is represented by a statistically significantly higher or lower percentage performance. Thus, we ameliorate otherwise persistent problems of arguing whether 70 percent, 85 percent, or 100 percent performance on a certain quality of care criterion is "good" care or not.

For some patients with a particular condition, the failure to comply with an applicable criterion can be a perfectly sound decision, not reflecting any decrement in the quality of care. For example, a blood sugar or urinalysis need not be ordered on *every* visit for a diabetic patient. The validity of our method depends on a pair of assumptions concerning how these exceptional cases are treated by practioners.

First, we assume that the populations of patients seen by each practitioner type are similar in their need for the criterion to be followed. If exceptional cases are not equally distributed among the practitioner types, then comparisons of compliance rates would reflect these different distributions and not necessarily any real differences in quality. Patient populations may also differ by the severity of the illness. Some criteria are more compelling for serious or complicated cases than for simple cases. Since at most clinics, triage of the patients sent more serious cases to the physicians more often than to the extenders, good quality would require *higher* compliance by the physicians with such criteria.

The second assumption pertains to the treatment of exceptional cases. Physicians and extenders may differ in their ability to recognize the cases that are exceptions to the criterion. Even if such cases are evenly distributed among practitioner types (i.e., if the first assumption holds), extenders may be more structured in their decisions, and may more often comply with the criterion in those cases where it is not strictly required. If extenders more often follow the criterion in cases where physicians would justifiably not, then compliance by extenders would be increased without any corresponding increase in quality. The increase in compliance might be sufficient to hide an otherwise poorer performance by the extender, or could even lead to the erroneous conclusion that the extender outperformed the physician. In considering the results from the analysis of performance on the criteria we have formulated, the likely frequency of exceptions to the criterion must be kept in mind. If they are very frequent, the effects of violations of these two assumptions could be important.

THE PATIENT CONTACT RECORD (ENCOUNTER FORM) AS A DATA COLLECTION INSTRUMENT

For all the measures described in following sections, data were collected on a Patient Contact Record (App. A); the Patient Contact Record was completed on all visits to primary medical care areas[2] of the four demonstration bases for one month during 1977.[3] Patients were asked to provide information about themselves and about the time it took to obtain an appointment if they had made one. Then, the health provider(s) furnished the medical details of the visit. The provider checked off the diagnosis or problem from a list of the more common diagnoses adapted from the ICDA classification system of the Royal College of General Practitioners, modified by the University of Rochester for use in training family practice

[1]These between-group comparisons must take statistical error into account. Appendix B discusses statistical techniques applied in these comparisons.

[2]These areas included: adult primary care clinic, flight surgeon's clinic, physical examination section, emergency room, and (if separate) internal medicine clinic.

[3]A more rudimentary version of the Patient Contact Record had been fielded in 1974 on visits to all outpatient clinics, at seven bases for two weeks, and at two bases for a six-month period.

physicians, and further modified by us. The list was found to account for 80 to 85 percent of the diagnoses in the outpatient clinics in our sample.[4] The categories for rating seriousness of condition were taken from the National Ambulatory Medical Care Survey. We devised our own lists of procedures, tests, and medicines.

Because of inherent limitations of the Patient Contact Record (which it shares with any other type of encounter form, and which are described in greater detail below), we confined ourselves to evaluating quality of care as manifested in the technical process of care (what is done for the patient). Within these bounds, we developed more than 60 quality of care criteria. Some apply to all persons with a certain diagnosis or problem, regardless of how many others were checked, while others apply only if the one diagnosis or problem was checked; some apply to every visit for the diagnosis or problem, while others apply on the first visit only.

It is worth noting in detail the strengths and limitations of this encounter form.

Its greatest advantages are that it is completed on the spot, and that the crucial information is furnished by the actual provider of care. We observed conscientious efforts by practitioners to complete the forms.

The form's limitations fall into two categories: limits on the breadth of information, and limits on its reliability. The reason for visit was not available,[5] nor did we have data concerning the results of laboratory tests or physical examination. In addition, a patient may have a combination of illnesses, or an illness of particular severity (so that a procedure justified for one patient would be inappropriate for another). Various factors of cultural, psychological, social, economic, or geographic importance may suggest the wisdom of treating a disease differently from the usual case, or the provider may be persuaded to deviate from intended care because of the patient's strong wish to omit or add some procedure. And, obviously, the provider is limited to the diagnostic and procedural categories provided for checkoff on the encounter form (these categories may not accurately describe some disease conditions or services rendered).

In regard to reliability, we must rely on the encounter form to convey a picture of what actually occurred at a particular visit, and we assume that this picture represents a reasonable estimate of the typical performance of these practitioners. However, the practitioner may have made the wrong diagnosis, and thus may have distorted our measurement by receiving credit for "correct" treatment of the wrong ailment. Or the provider may make a recording error when checking the diagnosis or procedure box. Such errors cause concern only if their incidence is systematically related to type of practitioner. If they occur randomly among practitioners, and are equally likely among the three practitioner types, the performance comparisons will be unaffected. Finally, practitioners' performance may be influenced by awareness of the ongoing study (the so-called "Hawthorne Effect").

Most of these limitations, rather than being specific to the Patient Contact Record, are generic to evaluation of quality of care, regardless of the techniques employed. In our opinion, none of the limitations, singly or in combination, outweighs the advantages of this instrument. However, these limitations do underscore the need to select simple, condition-specific criteria or other indicators of quality, and to interpret results cautiously.

SELECTION OF CONDITION-SPECIFIC CRITERIA

The criteria we selected are simple in design, so they can be easily applied to the encounter data. However, there are more important reasons for using "simple" instead of "sophisticated"

---

[4] *Appendix A also provides a frequency table of conditions marked on the Patient Contact Record by each type of practitioner.*

[5] *An attempt to collect reason for visit was made, but categories were too few and too broad to be useful.*

criteria. For one thing, they will apply to a larger number of patients. For another, simple criteria are more likely to gain acceptance among providers and therefore carry much greater weight if they reveal significant similarities or differences in the quality of medical care provided.

The authors accept responsibility for the medical content of all the criteria selected, but refer the reader to eleven sources of ambulatory quality-of-care criteria that they found useful in formulating the condition-specific criteria that were used. (See Bibliography, "References for Quality of Care Criteria.")

OTHER MEASUREMENTS OF QUALITY

Condition-specific criteria lie at the heart of this presentation, but we examined the quality of care with two other measures. The other measures are not criteria, strictly speaking, since we did not establish in advance normative statements of what the care should look like.

Ordering rates for selected tests and procedures, which we call utilization, can also be construed as a measure of quality, because unsuitable ordering can not only be wasteful of time, materials, and money but also may lower the quality of care. Utilization was our first additional measure of quality.

Utilization analysis examines the question of appropriate utilization of tests and procedures by comparing utilization rates for physicians, PAs, and PCNPs across a number of tests and procedures, including the "procedure" of ordering return visits. Comparison of ordering rates for each type of practitioner provides an objective approach to this question.[6] The results have implications for both cost and quality. If there are observed differences in ordering rates among groups of providers, the group ordering at a lower rate may be "underutilizing"; or, the group ordering at a higher rate may be burdening the system by "overordering".

It is also possible, however, that a high rate of ordering is a positive sign, in that it may really reflect a higher level of quality of care. We thus quickly reach an ill-defined zone where quality and utilization are virtually inseparable. It is difficult to know whether differences in ordering rates represent higher quality or lower quality, neglect or parsimonious prudence, waste or merely reasonable caution.

For both utilization and quality, utilization analysis can signal potential problems, identifying areas where chart review might permit more certain conclusions to be drawn concerning the presence of high or low quality, or too high or too low utilization. Appendix D provides a review of selected papers that attempt to deal with the middle zone where considerations of utilization and quality overlap.

After first presenting general data on the ordering of tests and procedures, Sec. IV below, on utilization analysis, probes more deeply into the variables that may influence the rate of ordering tests. One variable is the diagnostic mix: One type of provider may overorder a given test for one diagnostic group but, because this type of provider sees a smaller number of patients in the group, this type of provider may end up with the same gross ordering rate as another type of provider that orders appropriately for a larger number of patients in this group. Another variable is the ratio of initial visits to return visits for a particular problem. If that ratio of visits differs for physicians as opposed to PAs or PCNPs, the rate of ordering tests might well vary, since procedures are not ordered equally often on initial and return visits.

For the second additional measure of quality, we analyzed the amount and character of

[6]The statistical method used to compare ordering rates was the same method used for the condition-specific criteria and described in App. B.

physician supervision of cases handled by PEs. Rates of consultation with and referral to physicians were computed for various categories of visits. This analysis does not have physician performance available as a comparative yardstick; however, we have drawn comparisons with findings from another study of PE performance.

SAMPLE CONSIDERATIONS

The results presented in this report rely on analysis of a sample of visits collected at the four demonstration base hospitals. We believe that the sample is reasonably random and thus gives a reliable picture of the care provided. We also believe that the demonstration bases are representative of other Air Force hospitals of similar size, and that the results obtained in this study would apply to these other hospitals.

Clinic personnel collected data using the PCR on all visits for a four-week period. Table 1 shows the sample's time period and the number of records collected at each base in the primary medicine clinics (including family practice, general therapy, flight surgeon, physical exam, emergency room, walk-in, sick call, and general internal medicine clinics). If direct activity in the sample period is typical, it should be representative of the year as a whole; we have no evidence that it was not.[7]

We believe that the demonstration bases and personnel are reasonably representative of the Air Force Medical Service. The bases were chosen from a variety of commands and were typical of middle-sized Air Force hospitals. Bases were also selected that were likely to undergo as few transfers of personnel as possible, so as to establish a stable manning ratio for the demonstration. This meant that the demonstration bases already had fewer physicians and more PEs than the typical base hospital. New physicians were added to the hospitals in the course of normal rotation. New PEs were usually assigned from the newly graduated class of PAs, with no attempt to choose only the better graduates.

Table 1

BASIC DATA ON PRIMARY MEDICINE CLINICS RECORDED DURING SAMPLE PERIOD, 1977

Item	Chanute	Dyess	Fairchild	Nellis
Total number of MDs	12	8	15	13
Panel PAs	7	5	5	6
Panel PCNPs	1	2	2	2
Sample collection dates	3-30/4-26	3-17/4-13	4-27/5-27[a]	4-13/5-11
No. of patient visits	5010	4304	3780	4959

[a]Excludes May 3, 4, and 5, 1977, which were dedicated to health screening examinations.

Because of the relatively small number of PCNPs, however (see Table 1), conclusions about their performance cannot be confidently applied to PCNPs in the rest of the Air Force. The larger numbers of PAs and physicians give estimates of performance that could more reliably be expected to apply to the Air Force as a whole.

[7]During the sample period at Fairchild, three days of the clinic time were dedicated for health screening exams. These days were removed from the sample and the four-week period was extended an additional three days.

III. CONDITION-SPECIFIC CRITERIA

BACKGROUND

We first turn to an evaluation of quality of care through the use of condition-specific criteria. The policy issue is whether the Air Force can deliver the same quality of medical care when physician's assistants (PAs) and primary care nurse practitioners (PCNPs), working under the general supervision of a physician, provide a substantial portion of the outpatient care formerly provided by physicians.

This section compares the physician's extenders' (PEs') quality of care with that of physicians, for routine outpatient conditions that the PEs are trained to treat.[1]

We use the descriptive data collected on Patient Contact Records for a one-month period at the four demonstration bases. The data come from the group of selected clinics at each base that provide primary care adult medicine.[2]

SELECTION OF CRITERIA

Section II has discussed strengths and weaknesses of our data collection instrument, the Patient Contact Record. By concentrating on routine cases of low or moderate complexity (which account for most cases seen in Air Force outpatient clinics), we believe we avoided the difficulties and disagreements that sophisticated criteria probably would have engendered. Measures of technical process make sense in the outpatient setting, where many conditions being treated are common and conventionally managed, and where there is a relatively clear connection between process of care and outcome (cure or symptom relief). Finally, we used team physicians' compliance with the same criteria as a benchmark against which to judge the quality of care rendered by team extenders. The physicians were, of course, working in the same primary medicine settings.

CLASSES OF CRITERIA

We divided our criteria into five classes: desirable diagnostic actions, desirable therapeutic actions, undesirable diagnostic actions, undesirable therapeutic actions, and desirable disposition actions. Table 2 provides an example from each class, and shows the format in which we present the results of all our condition-specific criteria. The word "only" in the second, third, and fourth column heads signifies that we limited the application of our criterion to those visits where it was clear that only one MD, PA, or PCNP provided the care. (We thus excluded, for example, the rare visits when physicians and PAs together provided care.)

For each type of provider, Table 2 documents a percentage of compliance with the criterion and the number of visits to which the criterion was applied. (Note: If the criterion is specifically limited to first visits, the number of visits equals the number of patients involved; however, if the criterion is applied to all visits within a given diagnostic category, the number of patients is likely to be smaller than the number of visits. The *n* that we record is the number of visits.)

---

[1] See App. A for the complete list of conditions and the frequencies with which they were marked by each type of practitioner.

[2] Some of our analysis pools the data collected at the four bases, a reasonable procedure since there was no major variation from base to base. Exceptions to the finding of lack of significant variation are discussed in Apps. B and C.

Table 2

EXAMPLES OF REPRESENTATIVE CRITERIA AND COMPARATIVE MEASUREMENT OF COMPLIANCE

	Criterion Class	MD Only	PA Only	PCNP Only	Difference Statistically Significant?
I.	Desirable diagnostic action: Urinary tract infection (first visit); Urinalysis or urine culture ordered	90% n=19	75% n=69	90% n=20	no
II.	Desirable therapeutic action: Otitis media (infectious) (first visit); appropriate class of antibiotic prescribed	82% n=51	92% n=100	80% n=30	no
III.	Undesirable diagnostic action: Number of EKGs for "Non-EKG conditions" Total number of visits for "non-EKG conditions"	0.7% n=2581	0.4% n=5573	0.1% n=1343	MD > PCNP
IV.	Undesirable therapeutic action: Viral syndrome with gastroenteritis; antibiotic presecribed	1% n=86	2% n=164	9% n=58	PCNP > MD
V.	Desirable disposition action: Tranquilizer or antidepressant prescribed; definite followup planned	54% n=143	49% n=115	38% n=24	no

The last column of Table 2 indicates whether between-group differences were statistically significant at the $P < 0.05$ level, using the test for difference in proportions.[3] In other words, the chances of getting the measured difference or a larger difference between practitioner types would be 5 percent, if the practitioners were actually complying at equal levels. When there is a statistically significant difference, it is recorded by showing which group or groups of providers have a percentage compliance with the criterion significantly greater than the other(s). In the case of desirable actions, a higher percentage is consistent with better care. However, as will be emphasized in greater detail below, in the case of undesirable actions, a higher percentage represents worse care.

REPRESENTATIVE CRITERIA

The first criterion says that if the patient made a first visit for a problem diagnosed by the provider of care as urinary tract infection, the provider then ordered either a urinalysis or a urine culture.

Under desirable therapeutic action, the sample criterion reads: If the patient made a first visit which was diagnosed by the provider as infectious otitis media, then the provider prescribed an appropriate class of antibiotic (i.e., penicillin, erythromycin, or sulfa).

Through undesirable diagnostic actions, we can approach the question of "overordering" by various groups of providers. Note that because the action is undesirable, a lower percentage suggests a better ordering pattern. For example, are practitioners ordering electrocardiograms for a group of diagnoses where electrocardiograms are unlikely to have been needed?

Once again, with undesirable therapeutic actions, a lower percentage implies better care, because the action is undesirable. The representative criterion reads: If the provider selected the diagnosis of viral syndrome with gastroenteritis, then an antibiotic was prescribed. In this instance, because it is considered poor medicine to prescribe an antibiotic, a lower percentage compliance represents better care.

A group of desirable disposition actions constitutes our final criterion. The representative criterion reads: If a tranquilizer or antidepressant was prescribed, then a definite follow-up visit was planned.

BREADTH OF CRITERIA

It is of interest to know the "amount of care" upon which our technical-process, quality-of-care criteria impinge. Although each criterion views only one aspect of the care of a particular problem, and therefore does not thoroughly evaluate care for each condition, the full set of criteria taken as a whole does evaluate a substantial portion of care provided by each group of practitioners.

We have computed the percentages of all visits to team members for conditions specifically evaluated by our criteria.[4] The results are presented in Table 3. By "specifically evaluated," we mean that the condition was mentioned by name on our list of criteria. Thus, for two reasons, the reported percentage underestimates the total amount of care evaluated: first, the denominator includes routine examinations, whose content is fixed and to which criteria could not be applied; second, we did not add to the numerator the visits for various groups of

[3]This and other tests for statistical significance are discussed in App. B.

[4]The complete list of condition-specific criteria upon which we base our conclusions is found in Tables C.1 to C.5, App. C. We report on 62 criteria in all: 17 desirable diagnostic actions, 14 desirable therapeutic actions, 8 undesirable diagnostic actions, 16 undesirable therapeutic actions, and 7 desirable disposition actions.

conditions (for example, the "follow-up desirable" group of conditions—see the list of groups of conditions in Table C.6, App. C).

Table 3

AMOUNT OF CARE EVALUATED BY 1977 CRITERIA

Type of Provider	Number of Visits to Which Criteria Apply Divided by Total Number of Visits	% Visits Evaluated by the 1977 Criteria
Total visits to individual team members	5919/14,711	40.2%
Visits to MD only	1691/4429	38.2%
Visits to PA only	3280/8227	39.9%
Visits to PCNP only	948/2055	46.1%

Thus, depending on the mix of conditions seen by each provider, the criteria were applicable to between 38 percent and 46 percent of all visits seen individually by providers. For the group as a whole, the 1977 criteria were applicable to approximately 40 percent of all visits.[5]

SUMMARY COMPARISON

Having reviewed the five classes of quality of care criteria, we turn our attention to the summary comparison, which reflects the criteria listed in App. C. We first chose 42 of the total of 62 criteria in constructing the summary, and gave each criterion equal weight. A choice was made on which criterion to count when two or more criteria overlapped, and we attempted to use "best medical judgment" in making the choice. We then eliminated any criterion that had too few cases to be able to detect a true difference of 0.25 in compliance rates at least 75 percent of the time.[6]

The basic question was: How are PAs and PCNPs measuring up, insofar as we can determine with our criteria? Is their level of performance either equal to or better than the standard, i.e., compliance by team physicians with the same criteria? The results were striking. Very few significant differences between the provider groups emerged from the comparison. Directing our attention to the bottom line of Table 4, entitled "Total Actions", we see that, for 5 of the 28 criteria, PAs' performance was (statistically) significantly superior to that of physicians. For 20 of the 28, there was no difference between the performance of the PAs and the physicians. Therefore, for 25 out of 28 quality criteria, PAs' performance, on a statistical basis, equaled or exceeded the performance standard we set. PAs performed worse on 3 criteria out of 28.

Performance of the PCNPs was (statistically) significantly superior to that of physicians

[5]This percentage compares with the 20 to 25 percent of visits to which we were able to apply a more limited set of criteria in an earlier study of PA quality discussed below.

[6]This standard reflects the power of the statistical test. The calculation of each test's power is discussed in App. B.

Table 4

Summary Comparison: Quality of Care in Primary Medical Settings, Demonstration Bases, 1977

Criterion Class	PA Better	PA Equal to MD	MD Better	PCNP Better	PCNP Equal to MD	MD Better
Desirable diagnostic action	1	5	0	1	2	0
Desirable therapeutic action	2	3	1	1	0	1
Undesirable diagnostic action	0	3	1	1	2	1
Undesirable therapeutic action	1	7	0	0	7	1
Desirable disposition action	1	2	1	0	0	2
Total actions	5/28	20/28	3/28	3/19	11/19	5/19

NOTE: The significance tests for this table are based on the critical value for two contrasts. Tests using the single-contrast critical value for MD-PA comparisons give similar results: Based on total actions, PAs perform better on 6 out of 28 criteria and equal MD performance on 19 out of 28; MD performance excels PA performance on 3 out of 28.

for three of the 19 measures,[7] and was (statistically) the same for 11 of the 19 criteria. Therefore, for a total of 14 out of 19 quality criteria, PCNPs' performance equaled or exceeded the performance standard. PCNPs performed worse on 5 criteria out of 19.

Another measure of differences in overall performance can be calculated by using the sign test. Summarized in App. B, the results of this test show that PCNPs do not differ significantly from physicians; in 21 of 42 non-overlapping criteria, PCNPs outperform physicians. However, PAs' compliance with criteria exceeds physicians' compliance on 28 of the 42 criteria, a result unlikely to occur ($p < .05$) if PAs and physicians were performing at the same level.

Table 5 lists those criteria, from among those used in the summary comparison, on which PEs and physicians excelled each other.

Inspection of the small number of significant differences in performance reveals a feature that is medically reassuring: There is a lack of systematic deficiency on the part of any group of providers. That is, the few differences that there are, are scattered among the five classes of criteria. Upon comparison of physicians and PAs, the 3 criteria according to which physicians performed better are scattered among three classes; so are the 5 criteria according to which PAs performed better. For physicians and PCNPs, the 5 criteria according to which physicians performed better are scattered among four classes; the 3 criteria according to which PCNPs performed better are scattered among three classes. Because we believe that each criterion has medical merit (although the criteria are not all of equal importance or validity), it is impossible to say that any criterion showing a difference is "unimportant"; yet, the predominance of criteria showing equal or superior performance by PEs is the key finding, and the typical criterion showing similar or superior performance is certainly equal in "importance" to the typical criterion showing inferior performance.

Additional analysis leads to further insight. For example, significant differences in the performance of PAs and PCNPs occurred for only three of the criteria (two of which were used

[7]The number of criteria for PCNPs is less than that for PAs because more criteria were discarded because of insufficient power in the tests—the result of fewer PCNP visits.

Table 5

Criteria on Which PE and Physician Performance Differed Significantly: Primary Medical Settings, Demonstration Bases, 1977

MDs performed significantly better than PAs:

Urinary test infection (only diagnosis, first visit) → antibiotic prescribed.
Coryza (only diagnosis, first visit) → throat culture ordered. (Lower percentage is desirable.)
Condition that should usually be followed up → definite return appointment planned.

MDs performed significantly better than PCNPs:

Urinary tract infection (only diagnosis, first visit) → antibiotic prescribed.
Backache alone or with sciatica (only diagnosis, first visit) → x-ray ordered. (Lower percentage is desirable.)
Viral syndrome with gastroenteritis → antibiotic prescribed. (Lower percentage is desirable.)
Condition judged "very serious/serious" by provider → definite follow-up planned.
Condition that should usually be followed up → definite return appointment planned.

PAs performed significantly better than MDs:

Pharyngitis (only diagnosis, first visit) → throat culture ordered or penicillin prescribed.
Acute sinusitis (first visit) → any medication prescribed.
Coryza or febrile cold → antibiotic prescribed. (Lower percentage is desirable.)
Infectious otitis media (first visit) → decongestant prescribed.
Infectious otitis media (only diagnosis, first visit) → definite follow-up planned.

PCNPs performed significantly better than MDs:

Pharyngitis (only diagnosis, first visit) → throat culture ordered or penicillin prescribed.
Infectious otitis media (first visit) → decongestant prescribed.

$$\frac{\text{No. of EKGs on young adults without "EKG condition"}}{\text{No. of visits for other than "EKG condition"}}$$: Lower fraction is desirable.

in the summary comparison):

- Otitis media (*infectious, first visit*): antibiotic prescribed
- Backache (only diagnosis, first visit): x-ray ordered
- Condition judged "very serious/serious": definite followup planned

For all three of these criteria, PAs performed better.[8] For 59 other criteria, however, there was no significant difference in performance between these two groups of extenders. We therefore conclude that PAs and PCNPs perform about the same.

COMPARISON OF 1977 AND 1974 RESULTS

In 1974, using similar methods, we measured quality of care rendered by PAs.[9] Only a few of the criteria applied in 1977 had also been applied in 1974—a more limited Patient Contact Record in 1974 constrained the choice of criteria. Similar but not exactly the same criteria were formulated in 1974 for otitis media (a prescription needed on the first visit), acute sinusitis (a prescription needed on the first visit), urinary tract infection (a prescription needed on the first visit), and pneumonia (a chest x-ray needed on the first visit). For two other criteria, pharyngitis (throat culture needed on first visit), and diabetes (blood sugar needed on first visit), the criteria in 1974 and 1977 were virtually identical.

When 1977 results are compared with 1974 results (Table 6), there is no evidence of any worsening in PAs' performance—despite PA classes having been trained at different times, a number of PAs' having been out of school for a longer time, and a higher ratio of PAs to supervising physicians. This finding is encouraging.

Data were also available to explore further the question of differences in performance among various cohorts of PAs. (1974 data did not include information on PCNPs.) Analysis of individual performance on criteria can show whether differences in individual performance rates are related to any known characteristics of individual practitioners. One characteristic of interest is year of PAs' training. By comparing the performance of PAs who graduated from the Air Force's training program before June 1975 with those who graduated after June 1975, we can test whether the performance levels in 1977 were maintained, both by the cohort of early PA graduates and by the new cohorts of graduates entering the PA force.

Table 7 shows PAs' performance on a selection of criteria comparing two cohorts of PAs. The "New PA" group includes those who graduated from the training program after June 1975; the "Old PA" group consists of those trained before June 1975. There was little difference in their performance. Recent PA graduates perform as well as earlier graduates, as measured by compliance with these criteria. These thirteen criteria are broadly representative of the complete list of criteria. Criteria were chosen to show a range of PAs' performance—criteria where the PAs both exceed and fail to meet the physicians' performance levels. The sample sizes for each criterion had to be large enough to produce adequate statistical power, and criteria were also chosen to include selections from each of the five categories of criteria.

---

[8]The critical value ($d = 2.24$) used in these PA-PCNP comparisons was the same as used in the physician-PA and physician-PCNP comparisons, although the addition of another contrast (PA-PCNP) should require a higher critical value. Since we are primarily interested in the physician-versus-extender comparison, the critical value was chosen based on these two contrasts. The same value was used here to maintain comparability.

[9]N-1184-AF, "The Quality of Air Force Outpatient Care: How Well Do Physician Assistants Perform?"

Table 6

COMPARISON OF PA PERFORMANCE IN 1974 AND 1977

	Criterion	1974 Rate	1977 Rate	Statistically Significant?
1.	Diabetes Mellitus (first visit); blood sugar ordered (all visits included in 1977)	29% n=7	51% n=102	no
5.	Pharyngitis (only, first visit); throat culture ordered	70% n=155	84% n=223	1977 > 1974
7.	Urinary tract infection (only, first visit); urine culture or urinalysis ordered	79% n=77	75% n=69	no
12.	Pneumonia (first visit); chest x-ray ordered (sputum culture included in 1977)	67% n=12	80% n=20	no
19.	Acute sinusitis (first visit); any medication prescribed	93% n=55	98% n=60	no
23.	Otitis media (first visit); any medication prescribed (limited to antibiotics in 1977)	91% n=180	94% n=100	no
27.	Urinary tract infection (only, first visit); any medication prescribed	77% n=77	64% n=69	no

CONCLUSION

Despite the limitations of condition-specific measurements of quality, there is strong evidence that, for the kinds of conditions that they are trained to treat, both PAs and PCNPs perform at an acceptable level of quality, when physicians' compliance with the same quality criteria is employed as a standard of reference.[10] We find that the extenders deliver acceptable quality of care in this setting. Insofar as we can measure with our condition-specific criteria, PEs are safe for routine, outpatient conditions, in that they deliver a technical process of care equal to that of physicians.

[10]We wish to add that these objective results are consistent with the high subjective ratings given to these extenders by their supervisors.

Table 7

COMPARISON OF PA COHORTS GRADUATING BEFORE AND AFTER JUNE 1975

	Criterion	"Old PA"	"New PA"	Statistically Significant?
3.	Hypertension (only); blood pressure taken	56% n=168	65% n=182	Yes
6.	Pharyngitis (only, first visit); throat culture or penicillin prescribed	92% n=64	91% n=160	No
7.	Urinary tract infection (only, first visit); urine culture or urinanalysis ordered	70% n=27	79% n=42	No[a]
18.	Acne treated with antibiotic; antibiotic is tetracycline	72% n=36	80% n=81	No
24.	Otitis media (infectious; first visit); penicillin, erythromycin, or sulfa	90% n=40	93% n=61	No
26.	Otitis media (noninfectious, first visit); decongestant prescribed	83% n=24	92% n=108	No
28.	Urinary tract infection (only, first visit); antibiotic prescribed	37% n=27	45% n=42	No[a]
36.	Backache alone or with sciatica (only, first visit); x-ray ordered	8% n=25	17% n=47	No[a]
39.	Coryza (only, first visit); throat culture ordered	49% n=85	51% n=217	No
41.	*Coryza or febrile cold*; antibiotic prescribed	9% n=241	10% n=465	No
42.	Viral syndrome without gastro-enteritis; antibiotic prescribed	14% n=105	15% n=94	No
51.	No. of times oral or injected steroid was prescribed, divided by no. of visits for "steroid-contraindicated conditions"	0.2% n=433	0.2% n=527	No
56.	Condition judged very serious/ serious by provider; definite follow-up planned	87% n=55	78% n=206	No

[a]The power of these tests failed to meet the criteria of at least 75-percent chance of detecting a difference of 0.25 in the true compliance rates.

IV. UTILIZATION ANALYSIS

INTRODUCTION

Appropriate ordering of tests and procedures is another important facet of good quality medical care. A portion of the previous section examined undesirable diagnostic actions and undesirable therapeutic actions by selecting specific tests, procedures, or prescriptions (e.g., electrocardiogram, physical therapy, steroids) and analyzing whether or not they were employed in situations where they were likely to have been inappropriate.

This section adopts a different approach: "utilization analysis." As explained in Sec. II, results of utilization analysis can have implications for both cost and quality. However, we pay particular attention to the question of higher ordering rates as a sign of possible overutilization, because of the often expressed fear that physician's extenders (PEs) might overburden a system by overordering. Although we raise the issue of quality from time to time as we review the data, we give relatively more weight to considerations of cost.

This section presents data on the ordering of tests and procedures—first in general, and then for specific diagnoses and conditions. It next examines the pattern of "disposition," the decision made at the end of every visit as to what kind of follow-up will be planned. Disposition is treated in general only, because the observed pattern is clear and explicable. Finally, we reach the conclusion that no serious weaknesses are discernible in the utilization patterns of PEs, and we briefly discuss the implications of this finding.

METHODS

The data collection methods were described in Sec. II, but to repeat: Data collected on Patient Contact Records during the 1977 survey were used. Data are presented for patients seen only by a physician, or by a physician's assistant (PA), or by a primary care nurse practitioner (PCNP). We have excluded tests ordered during the infrequent visits when a patient saw either none or more than one of these three types of providers. For all tables, statistical significance is tested at a level of $p < 0.05$. That is, we say that the difference between two practitioners is significant when the chance of measuring such a difference occurring in a sample of visits is less than 5 percent if the two practitioners were actually ordering at the same rate. We examine the appropriateness of utilization by computing the number of times a specific test or procedure was ordered during 100 visits for all conditions, or for a selected group of diagnoses.[1] Although it might be unwise to make value judgments concerning the suitable number of times that ordering should occur, we would be troubled if physicians, PAs, and PCNPs were ordering at widely varying rates. These rates are also helpful to the administrator who wants to predict the number of tests that he can expect the staff to order.

ORDERING FOR ALL CONDITIONS

We computed the rate of ordering for the following tests or procedures: any type of x-ray, physical therapy, electrocardiogram, urinalysis, electrolytes, and blood counts. (The entire set

[1] Table C.6 in App. C specifies the diagnoses that comprise each group.

of visits was considered, not merely the 40 percent of visits covered by the criteria in the previous section.) Table 8 shows the gross rate of ordering these tests or procedures per 100 visits for patients seen only by a physician, or only by a PA, or only by a PCNP. Upon review of the rates, it is apparent that there are no large absolute differences in rates of ordering, for all problems taken as a group. For example, with the blood count category, we find that PAs are ordering approximately one-half count more per 100 visits more than are physicians or PCNPs. This difference would not seem to indicate misordering: It is not statistically significant and is unlikely to be clinically or economically significant, either. With urinalysis, the PAs are ordering approximately one and a half more urinalyses per 100 visits than are physicians. Even though this difference is significant, it is doubtful that any clinical or economic significance can be attached to it.

Table 8

RATES OF ORDERING TESTS PER 100 VISITS, FOR ALL CONDITIONS

Test	MD n=4429	PA n=8227	PCNP n=2055	Statistically Significant?
Any x-ray	9.60	*11.6*	11.1	yes
Physical therapy	0.23	*1.51*	*1.46*	yes
Electrocardiogram	*3.66*	1.65	1.22	yes
Urinalysis	6.05	*7.72*	7.30	yes
Electrolytes	2.91	2.80	3.75	no
Complete blood count, white blood count, hematocrit	7.42	8.00	7.49	no

Two rather marked differences are apparent, and they occur in opposite directions: PAs are ordering approximately two more x-rays per 100 patient visits than are physicians, while physicians are ordering approximately two more electrocardiograms per 100 visits than are PAs. In both cases, the ordering rate for PCNPs is quite close to that of PAs. Both of these findings are statistically significant. One could conjecture about possible clinical significance (and we will do so below), but the economic importance of these differences is likely to be minor because of the low marginal cost of additional x-rays and electrocardiograms in Air Force facilities.

Taking the categories together, we note that for three of these tests, PAs order significantly more tests; for one of them, PNCPs order significantly more; and for another, physicians order significantly more. Throughout this discussion, tests of statistical significance are fashioned in order to compare the two types of PEs with physicians, and not with each other.

ORDERING FOR SPECIFIC CONDITIONS

The gross calculations above do not probe the variables that may influence the ordering of tests. They do not really tell us if PAs or PCNPs are overutilizing tests and procedures in comparison with physicians. Only when variables such as diagnostic mix and the proportion of initial visits to return visits are taken into account can we perform a more exacting analysis of different rates of ordering tests. Even then, the reader will observe that we are left with the question of the reason for any observed differences in ordering rates: We reach that ill-defined zone where quality and utilization are inseparable. Nevertheless, we acquire useful informa-

tion on utilization of services, and we signal potential problem areas relating to both utilization and quality.

The tables that follow specify a narrower range of diseases, in order to lessen the effect of diagnostic mix on the results. Likewise, they separate first visits from return visits.

ORDERING OF X-RAYS

Turning to disease-specific rates of ordering tests, we start with x-rays (Table 9). The top section of Table 9 controls for disease categories. "X-ray conditions" is the general term for a group of disorders, all of which might well require x-rays at some time for some proportion of patients being seen. The five more specific disease conditions listed (i.e., headaches, ischemic and other heart diseases, acute and chronic sinusitis, backache, and arthritis and joint pain) are all specific subsets of the general term "x-ray conditions." This table limits the count of visits to those encounters for which the specific diagnosis was the sole one written on the Patient Contact Record.

Now that we have taken case mix into account, we notice dramatic changes in the numbers. Before (Table 8), both PAs and PCNPs had ordered more x-rays than physicians per 100 visits for all conditions. We now note an even larger difference (and a statistically significant one) in ordering rate when we look at the more specific groups of diagnoses. In every case for PAs, and in every case but one for PCNPs, these two types of PEs are ordering more x-rays than are physicians for each of the diagnostic subsets; and the PCNPs are ordering significantly more x-rays for backache.[2] Is this overutilization, or is it higher quality of care?

The remainder of Table 9 controls for definite first visit versus definite return visit.[3] When we look at the differences in ordering rates for first visits and return visits, we observe that for "x-ray conditions" in general, the ordering rate is higher for first visits than for return visits. The significantly greater number of x-rays ordered on first visits for "x-ray conditions" by PAs and PCNPs does not extend to return visits, where there is no significant difference.

Despite the small number of observations, it is of interest that all three provider groups decrease their x-ray ordering rate for headaches on return visits. In the case of ischemic and other heart diseases, the physicians initially have a lower ordering rate, and their rate increases on return visits, while the PEs' ordering rates decline. In the case of sinusitis, for all provider groups, there is no trend toward a decreased ordering rate on return visits, which suggests that providers may become more concerned with some patients who have not progressed satisfactorily by the time of a return visit.

PCNPs show a marked and significantly higher rate of ordering x-rays for backache, a difference which is limited to the first visit. For backache, the PCNPs' ordering rate of more than 42 x-rays per 100 visits on first visit is so high that one might well question whether all of these x-rays were necessary. In the case of return visits for arthritis and joint pain, where both the PAs and PCNPs are ordering x-rays significantly more often than are physicians, it is impossible to guess if the patients really required x-rays. Overall (first and return visits), there is no significant difference in the ordering rate of x-rays for arthritis and joint pain among the three provider groups.

---

[2] For many diagnostic subsets (here in the x-ray section, and in following sections reporting other tests), numbers of cases are so small that it is difficult to demonstrate statistically significant differences. Yet, it is still possible to detect a clear pattern of higher utilization of x-rays (for appropriate conditions) by PEs—a pattern supported by the statistically significant differences found in the aggregate "x-ray conditions" category.

[3] The total number of visits is lower than in previous x-ray data because we have discarded all visits with ambiguous or unindicated first or return status. We include only those visits which we are certain are first visits or return visits.

Table 9

X-Rays per 100 Visits

Condition	MD	PA	PCNP	Statistically Significant?
First and Return Visits				
"X-ray conditions"	17.1 n=684	*24.7* n=1419	*22.4* n=317	yes
Headaches	4.3 n=46	12.7 n=102	16.7 n=24	no
Ischemic and other heart diseases	14.4 n=97	18.8 n=32	7.7 n=13	no
Sinusitis	14.3 n=49	27.6 n=116	24.0 n=25	no
Backache	8.6 n=58	14.1 n=185	*27.6* n=58	yes
Arthritis and joint pain	21.5 n=79	29.2 n=240	30.0 n=40	no
First Visits Only				
"X-ray conditions"	21.0 n=271	*31.6* n=686	*30.5* n=164	yes
Headaches	10.5 n=19	19.5 n=41	33.3 n=9	no
Ischemic and other heart diseases	6.3 n=16	27.3 n=11	33.3 n=3	no
Sinusitis	5.9 n=17	25.0 n=56	27.8 n=18	no
Backache	3.4 n=29	13.3 n=75	*42.9* n=28	yes
Arthritis and joint pain	48.0 n=25	42.9 n=112	35.3 n=17	no
Return Visits Only				
"X-ray conditions"	12.1 n=224	15.8 n=524	15.1 n=126	no
Headaches	0 n=16	7.1 n=42	10.0 n=10	--
Ischemic and other heart diseases	15.7 n=51	13.3 n=15	0 n=8	no
Sinusitis	20.0 n=10	29.2 n=48	20.0 n=5	no
Backache	18.8 n=16	14.8 n=88	11.1 n=27	no
Arthritis and joint pain	3.2 n=31	*20.7* n=92	*26.1* n=23	yes

ORDERING OF PHYSICAL THERAPY

Turning now to a procedure (physical therapy, Table 10), we note that the two PE groups order more physical therapy for a group of conditions where physical therapy might well be helpful. The difference is statistically significant; physicians are ordering far less physical therapy for those conditions. The three subgroups selected for physical therapy—physical therapy group (excluding fractures, sprains, and strains), arthritis and joint pain, and backache—also show higher ordering rates for PAs and PCNPs. These differences persist when return visits are viewed separately.

However, it seems clear that PEs are not overloading the physical therapy department. Referring back to Table 8, note that the difference in the rate of ordering physical therapy per 100 visits for all conditions amounts to fewer than two procedures ordered per 100 visits. It is possible, however, that the physicians are not ordering as much physical therapy as might be consistent with good quality (for example, no physical therapy ordered for patients with arthritis and joint pain or with backache). If there were reason to suspect that this were the case, it would be possible to review a group of charts bearing specific diagnoses to decide if a problem existed.

ORDERING OF ELECTROCARDIOGRAMS

Table 11 presents more detailed consideration of electrocardiograms per 100 visits. There are no statistically significant differences in this table. For the electrocardiogram group as a whole (first and return visits), physicians are ordering more tests than are the PEs. Reference to the lower parts of the table shows that this difference is accounted for mainly by ordering activity during first visits. The single subset entitled "ischemic and other heart diseases" contains too small a number of visits handled by PAs and PCNPs to draw any conclusions with confidence. It is certainly reasonable to conclude that electrocardiograms are not being overordered by PAs and PCNPs.

ORDERING OF URINALYSES AND OTHER TESTS

Table 12 shows the differences in rates of ordering urinalysis per 100 visits. There are no significant differences, whether the analysis is limited to urinary tract infection, or to a somewhat broader category of urinary tract infection along with urethritis and other urinary tract diseases. Physicians evidently see a smaller number of these cases than do PEs. This accounts for the statistically significant difference in ordering rates that appeared in Table 8.

The final two test categories listed in Table 8 were not analyzed by subsets of conditions, or by first and return visits. For all provider groups, the differences in ordering rates for these tests per 100 visits for all conditions were not statistically significant.

RATES OF ORDERING TESTS: APPLICATION

Even though it is difficult to determine whether the differences are attributable to variations in quality of care or patterns of utilization, utilization analysis is important because it signals potential problem areas. If desired, it would be possible to follow up, by means of chart review, any significant or inexplicable differences discovered. This method can therefore serve as a flagging device for uncovering areas where chart review might permit more certain

Table 10

PHYSICAL THERAPY PER 100 VISITS

Condition	MD	PA	PCNP	Statistically Significant?
First and Return Visits				
"Physical therapy group"	1.0 n=384	*9.1* n=963	*9.8* n=204	yes
PT group excluding fracture, sprain, strain	0 n=160	*9.3* n=525	*12.1* n=116	yes
Arthritis and joint pain	0 n=79	5.0 n=240	5.0 n=40	no
Backache	0 n=58	*14.1* n=185	*15.5* n=58	yes
First Visits Only				
"Physical therapy group"	0.6 n=176	*9.7* n=485	*11.3* n=106	yes
PT group excluding fracture, sprain, strain	*0* n=64	*9.0* n=244	*13.8* n=58	yes
Arthritis and joint pain	0 n=25	5.4 n=112	5.9 n=17	--
Backache	0 n=29	13.3 n=75	14.3 n=28	no
Return Visits Only				
"Physical therapy group"	1.0 n=102	*9.3* n=335	*8.4* n=83	yes
PT group excluding fracture, sprain, strain	0 n=54	*10.5* n=210	10.9 n=55	yes
Arthritis and joint pain	0 n=31	5.4 n=92	4.3 n=23	--
Backache	0 n=16	14.8 n=88	18.5 n=27	no

Table 11

Electrocardiograms per 100 Visits

Condition	MD	PA	PCNP	Statistically Significant?
First and Return Visits				
"Electrocardiogram group"	12.5 n=289	8.0 n=511	8.5 n=117	no
Ischemic and other heart diseases	25.8 n=97	21.9 n=32	7.7 n=13	no
First Visits Only				
"Electrocardiogram group"	21.1 n=38	12.5 n=96	15.0 n=20	no
Ischemic and other heart diseases	18.8 n=16	27.3 n=11	33.3 n=3	no
Return Visits Only				
"Electrocardiogram group"	8.9 n=191	7.5 n=333	7.7 n=78	no
Ischemic and other heart diseases	25.5 n=51	13.3 n=15	0 n=8	--

Table 12

Urinalyses per 100 Visits

Condition	MD	PA	PCNP	Statistically Significant?
First Visits Only				
Urinary tract infection (UTI)	60.0 n=45	63.9 n=155	80.0 n=40	no
UTI, urethritis, and other urinary tract diseases	43.9 n=82	54.9 n=255	60.9 n=69	no
First Visits Only				
UTI	84.2 n=19	71.2 n=69	95.0 n=20	no
UTI, urethritis, and other urinary tract diseases	55.6 n=36	61.8 n=110	71.0 n=31	no
Return Visits Only				
UTI	37.5 n=8	58.0 n=69	58.8 n=17	no
UTI, urethritis, and other urinary tract diseases	40.0 n=20	52.1 n=117	47.1 n=34	no

conclusions to be drawn concerning the presence of high or low quality, or too high or too low utilization of services.[4]

DISPOSITION IN GENERAL

In addition to the possible quality-of-care implications of overordering tests and procedures, such overordering (which has *not* been demonstrated here) would stress finances and personnel. A system could also be stressed by "overordering" of return visits to clinics. In medical jargon, the phrase "disposition of visit" refers to the provider's decision at the end of every visit on what will be the subsequent contact, if any, with the patient. We will review data for the end of first visits alone, then return visits alone, and finally for all visits combined. All tables show the disposition categories used by the National Ambulatory Medical Care Survey in its publications. The different possibilities for disposition range from no expected future contact for this problem, through telephone contact, to definite reappointment or referral. Admission to the hospital is also a possible disposition.

It is worthwhile to compare the dispositional behavior of various groups of providers. If one group orders many more return visits or referrals for its patient population than does another group with a similar patient population, it is likely that either one group is overordering or the other is underordering. Either possibility might reflect differences in the quality of care delivered; in addition, "extra" visits would, in and of themselves, stress the medical care delivery system, and would likely generate additional tests and procedures. Do our data in fact show overburdening or underutilization of the system by one or another type of provider? What happened in a situation in which the PEs had great freedom in deciding whether or not to make return appointments?

Table 13 provides this information for first visits only and for return visits only, while Table 14 provides it for all visits combined. Both types of visits, of course, necessitate decisions regarding follow-up.[5]

At the end of first visits, where, in general, triage was purposely not carried out, we see that disposition patterns are very similar among the three provider types. The "no follow-up" category reflects the acute and uncomplicated nature of many problems treated in the primary medical setting. There is little "telephone follow-up" by any type of provider. Definite return appointments are shown in the row labeled "return at specified time." These rates of definite return are virtually identical, which is not surprising given the similarity in patient population seen on first visits. This result provides satisfying support for the position that PAs and PCNPs are neither overordering nor underordering visits. This row also furnishes evidence that neither PAs nor PCNPs are generating an untoward number of return visits in comparison with physicians. Referral patterns are virtually identical as well.[6]

Testing for statistical significance using a Chi-square test, we find that the physician distribution of dispositions is significantly different from both the PAs' distribution (X^2 = 78.47, 5 d.f.) and the PCNPs' distribution (X^2 = 51.71, 5 d.f.). Because each category of disposition contributes to the total Chi-square statistic, the larger the contribution of any category

---

[4]The reader is invited to review the results of the other method we presented for investigating possible inappropriate utilization (i.e., selecting specific tests, procedures, or prescriptions and analyzing whether or not they were employed in situations likely to have been inappropriate). See App. C, "Discussion of Selected Criteria of Quality."

[5]All three tables include all physicians employed in primary care settings, not only those who are team members.

[6]*A priori*, one might expect a higher return and referral rate from PEs at the end of first visits, because certain patients who really should have been seen by a physician would have been "mismatched" to a PA or PCNP and would need to be reassigned to a physician for an additional visit. Our data do not demonstrate this pattern, probably because that type of patient usually has a problem that a physician could not deal with in a single visit, either. Therefore, the physician would also call for a definite return or referral.

Table 13

Disposition of Visits in Primary Medical Settings: Demonstration Bases, 1977

Disposition	MD Only	PA Only[a]	PCNP Only[a]
First Visits			
No follow-up/return if needed	58%	62%	63%
Telephone follow-up	2	<1	2
Return at specified time[b]	25	27	25
Refer to other MD	9	8	10
Return to referring MD	<1	<1	<1
Admit to hospital	5	2	<1
Total = 5383 visits	n=1449	n=3518	n=416
Return Visits			
No follow-up/return if needed	33%	47%[c]	51%[c]
Telephone follow-up	2	<1	2
Return at specified time[b]	54	41	34
Refer to other MD	9	10	12
Return to referring MD	<1	<1	0
Admit to hospital	2	1	<1
Total = 4679 visits	n=1399	n=2589	n=691

[a] PA and PCNP distributions are statistically significantly different from MD distributions ($p < 0.005$) using Chi-square tests (PA $X^2 = 78.47$, PCNP $X^2 = 51.71$, each with 5 d.f.).

[b] Includes 1.0 percent "Admit to Quarters."

[c] Again statistically different ($p < 0.005$, PA $X^2 = 95.81$, PCNP $X^2 = 96.11$)

Table 14

Disposition of Visits in Primary Medical Settings, Demonstration Bases, 1977: First and Return Visits Combined

Disposition	MD Only	PA Only[a]	PCNP Only[a]	Combined AF Providers	GP/FP, NAMCS[b]
No follow-up/return if needed	48%	57%	58%	54%	45%
Telephone follow-up	2	<1	2	2	4
Return at specified time[c]	35	32	29	32	51
Refer to other MD	9	9	11	9	3
Return to referring MD	<1	<1	<1	<1	1
Admit to hospital	4	2	<1	2	1
Total = 14063 visits	n=3810	n=7095	n=1956		

[a] PA and PCNP distributions are statistically significantly different from the MD distribution ($p < 0.005$) using Chi-square tests (PA $X^2 = 140.57$, PCNP $X^2 = 98.16$, each with 5 d.f.).

[b] 1975 data (Advance data, No. 15, December 14, 1977, p. 9).

[c] Includes 1.0 percent "Admit to Quarters."

to the statistic, the more that category helps to cause the conclusion that differences are significant. The difference between MDs and PAs in the use of telephone follow-up accounted for much of the significance of the Chi-square statistic. The differences in rates of hospitalization between physicians and both PAs and PCNPs also accounted for a large portion of the total difference.

An interesting finding is the distinctly higher hospitalization rate by physicians at the end of first visits, and the difference is maintained (though to a lesser extent) at the end of return visits. This finding suggests the presence of an informal triage system whereby patients who appear to be seriously ill are identified and steered to the physician even on the first visit. In fact, we observed such systems in action at the front desks of the teams.

Turning to return visits, we again note evidence of similar dispositional patterns and similar ordering of return visits on the part of all three provider groups. As expected, fewer of the patients who already have been told to come back for one return visit are told that no further follow-up will be needed. To the contrary, the row labeled "return at specified time"—that is, definite return appointments—is higher for all three groups on return visits as opposed to first visits. This change is consistent with the notion that a number of patients coming for return visits have chronic conditions that will require regular contact. In our opinion, the observation that physicians showed the highest rate of ordering definite return appointments, and the lowest rate for deciding that no follow-up is needed, does not imply that MDs overorder return visits or that PEs underorder; rather, we believe it reflects the difference in patient mix on return visits, when physicians are seeing a higher proportion of the more serious and more chronic cases.[7]

At the end of return visits, PAs and PCNPs both show insignificantly higher rates of referral. Allowing for the likelihood that the seriousness[8] of cases seen on return visits by PAs and PCNPs will be somewhat lower than the seriousness of cases seen by physicians, we conclude that when case mix is considered, PAs and PCNPs do generate slightly higher referral rates than do physicians. We do not find these referral rates worrisome. Had they been much higher than that of the physicians, we might have suspected overburdening of the system; at this rate of referral, however, it is most reasonable to conclude that the less highly trained providers are simply more cautious. There is certainly no gross imposition of visits on the system because of this extra caution.

Again, tests for statistical significance show differences between MDs' distributions and that of both PAs and PCNPs (PA $X^2 = 95.81$, PCNP $X^2 = 96.11$ with 5 d.f.). Greater frequency of return visits on the part of physicians and greater frequency of "no definite follow-up" on the part of PAs and PCNPs contributed the main portion of the significant Chi-square statistic. Lower PCNP rates for hospitalization were also important.

Table 14 above combines first and return visits. We include it to enable a comparison with the civilian sector. For purposes of comparison, we have combined all three Air Force provider groups. We note a higher rate of scheduled return appointments ("return at specified time") in the civilian sector (GP/FP NAMCS column). The difference might be attributable to differences in age distribution or complexity of problem in the two settings. There are more internal referrals within the Air Force system. However, when we add the Air Force's internal referrals to its definite return appointments, we still note more scheduled returns in the civilian sector. To speculate on the reasons for these differences would go beyond what our data permit.

---

[7]Several criteria presented in Sec. III in fact showed that physicians tend to order return visits more often for a group of conditions that would probably warrant a return visit.

[8]See Sec. V for further discussion of seriousness.

CONCLUSIONS

To summarize our analysis of utilization, let us review the principal findings.

First, the condition-specific criteria in Sec. III, which dealt with possible unwarranted ordering of tests, procedures, or therapies for specific conditions (the so-called "negative criteria") failed to demonstrate unnecessary utilization on the part of PAs and PCNPs as compared with physicians.

Second, our analysis of ordering rates for selected tests and procedures did not give evidence of misordering on the part of these two types of PEs.

Third, the documentation of disposition patterns shows that, with regard to internal generation of clinic utilization, all three provider groups behaved similarly. In short, there is no evidence for overordering or overutilization on the part of PAs and PCNPs.

V. SUPERVISION OF PHYSICIAN'S EXTENDERS

Both physician's assistants (PAs) and primary care nurse practitioners (PCNPs) are trained to work under the supervision of a physician. In a system such as that used at the demonstration bases, without structured triage, a physician must be available to supervise the extenders' handling of complex problems and to assume responsibility for the most serious cases. The availability of supervision does not in itself guarantee that the supervision is effective or that it improves the quality of care given, but the absence of the supervisory channel or lack of evidence of its use would signal potential problems.

While too little contact between extenders and supervisors could cause concern, too much supervision could also signal a problem. PAs and PCNPs are trained to handle a large number of routine problems in an outpatient setting without a physician's direct involvement. Supervision of an excessively large number of visits wastes physicians' time and suggests poor extender performance.

To investigate the adequacy of supervision during the demonstration project, we again analyzed the Patient Contact Record data. We wanted to know to what extent a physician was involved in the care of patients seen by an extender.

Several levels of involvement are possible. The extender may see that a patient's problems require a physician's expertise, and so *transfer* the patient to the supervising physician, or *refer* the patient to a specialist within the base hospital. The extender may *consult* the supervising physician, who will decide whether or not to speak with or examine the patient before advising the extender. Finally, the extender may need only a *countersignature* for prescriptions, referrals outside the hospital, or other specified orders or dispositions.

Even when the visit itself does not include any physician's involvement, a return visit by the patient might. Patients told to return to the clinic might be scheduled to see the physician team member on their return visit. Any patient admitted to the hospital will be reviewed by a physician.

Physician involvement ought to be more likely in cases where the patient has a problem that is complex or has serious potential consequences. It also ought to be more likely to occur (for a wider range of problems) if the PE believes the patient's condition to be serious or very serious.

We have divided visits to extenders into four categories of varying physician involvement. The first category, "consultation with physician," includes all visits where an extender sought some assistance from the supervising physician. Assistance may include seeing the patient, but it does not include cases where only countersignature was sought. The second category, "referred to physician," includes referrals to other clinics or CHAMPUS on the same day or for a future appointment, and also includes hospital admissions. The third category, "to return," includes all visits when the PE handled the case without a physician's involvement and when the PEs instructed the patient to make a return appointment, as well as visits when the patient was admitted to quarters. Since we cannot determine whether a physician or an extender saw the patient on the return visit, this category represents only the possibility of physician involvement. Finally, a fourth category includes visits when no physician was involved and when the patient was instructed to return or telephone if needed, or when no formal follow-up was planned.

We further divided the conditions listed on the Patient Contact Record into three groups: 1) physician generally not needed; 2) physician may or may not be needed; and 3) physician

generally needed. The first group included conditions where the PEs could treat the typical case of that condition without the need for direct physician involvement. While there may be some disagreement about what conditions extenders should be treating, this group includes conditions about which there is little question. The second group includes those conditions about which there might be a question: Some observers would be comfortable allowing extenders to treat these conditions without direct physician involvement, while others would prefer that the physician handle these conditions. Finally, the third group includes conditions where most would agree that a physician is needed.[1] This group comprises the few, most serious, conditions, e.g., malignant neoplasms (cancers) or arrhythmias (irregularities of heartbeat). For all three groups, the division was made on the basis of the diagnostic labels alone, with necessary disregard for particulars of the disease, or of the individual's specific medical status. Thus, for some "MD not needed" conditions, a consultation might well be appropriate, in certain circumstances. Similarly, some patients with conditions (e.g., chronic, stable atrial fibrillation—an arrhythmia) that must be designated as "MD needed" because of their diagnostic labels can in fact be appropriately managed by an extender without need for direct physician involvement.

Table 15 shows the extent of physician involvement in visits seen by PAs. The proportion of visits in which a physician actually participated is shown for each group of conditions. First visits and return visits are separated. (Where more than one condition is indicated for a visit, the condition most likely to need physician involvement is the one identified with the visit.)

Table 15 shows that the bulk of first visits seen by the PAs are for conditions usually not requiring physician involvement. For these visits, PAs rarely seek a physician's advice in handling the case. Only 3 percent of these cases involve a physician in the care, and only 8 percent are referred to a physician. Conditions that are more likely to require physician involvement are fewer in number. For these cases, PAs are much more likely to consult with a physician about the case (18 percent when "usually needed" and 12 percent when "may be needed"). PAs are also much more likely to refer these cases to a physician (22 percent when "usually needed" and 14 percent when "may be needed"). For return visits, the results are quite similar. Table 16 shows the same statistics for PCNPs. The level of physician involvement for PCNPs is similar to that for PAs.

Tables 17 and 18 show consultation rates for visits divided according to the extenders' evaluation of the seriousness of the problem. Again, we see a greater likelihood of consulting a physician for serious problems. For PAs the likelihood of seeking consultation increases twelve-fold between "not serious" and "very serious/serious" cases. PCNPs also are much more likely to seek a physician's advice on serious or very serious cases.

It is difficult to assess the appropriateness of the level of physician supervision and consultation that we see from these data. The rates of 5 percent for first visits and 8 percent for return visits do not put any serious burden on physicians' time. But is this level of physician supervision enough to ensure high-quality PE performance?

One standard for comparison comes from a recent evaluation study of PAs' performance in a prepaid health plan setting (Kaiser-Permanente Program, Portland, Oregon), a setting similar to the Air Force clinics in the demonstration (Record et al., 1977a). The study reviewed charts of patients who had seen PAs, and recorded the frequency and appropriateness of consultation in about 12 percent of the cases. These cases, however, included an unspecified number of requests for countersignature, a type of consultation excluded from our results for the Air Force clinics. A separate sample of consultations in the same clinic showed that 31 out of 83 consultations were for physician signature on forms or prescriptions. Applying this rate to the 12 percent average gives a rate of 7 to 8 percent for consultations excluding those for

[1]Appendix E lists the diagnoses specific to each group.

Table 15

PAs' use of Physician Consultation, Referral, and Return Appointments, by Main Condition's Need for a Physician

Extent of MD Involvement	MD Usually Needed	MD May Be Needed	MD Usually Not Needed	All Visits
	First Visits			
Consultation	18%	12%	3%	5%
Referred to MD	22	14	8	9
To return	28	32	24	25
No definite follow-up	32	42	65	61
Total	100%	100%	100%	100%
	n=40	n=628	n=2961	n=3629
	Return Visits			
Consultation	16%	10%	7%	8%
Referred to MD	14	9	9	9
To return	44	53	31	36
No definite follow-up	26	28	52	47
Total	100%	100%	100%	100%
	n=43	n=516	n=2188	n=2747

NOTES: An independent judge determined the need for a physician, based on main condition. Cases were eliminated if they provided incomplete data on the particular questions used for this table.

Table 16

PCNPs' use of Physician Consultation, Referral, and Return Appointments, by Main Condition's Need for a Physician

Extent of MD Involvement	MD Usually Needed	MD May Be Needed	MD Usually Not Needed	All Visits
	First Visits			
Consultation	11%	13%	5%	7%
Referred to MD	33	13	6	8
To return	33	30	21	22
No definite follow-up	22	44	68	63
Total	100%	100%	100%	100%
	n=18	n=179	n=853	n=1050
	Return Visits			
Consultation	0%	9%	9%	9%
Referred to MD	50	10	9	10
To return	30	48	26	30
No definite follow-up	20	33	55	51
Total	100%	100%	100%	100%
	n=10	n=128	n=553	n=691

NOTES: An independent judge determined the need for a physician, based on main condition. Cases were eliminated if they provided incomplete data on the particular questions used for this table.

Table 17

PAs' use of Physician Consultation, Referral, and Return Appointments, by Seriousness of Condition Treated

Extent of MD Involvement	Very Serious/ Serious	Slightly Serious	Not Serious
First Visits			
Consultation	36%	10%	3%
Referred to MD	17	16	7
To return	35	49	19
No definite follow-up	12	25	71
Total	100%	100%	100%
	n=83	n=650	n=2974
Return Visits			
Consultation	32%	12%	4%
Referred to MD	10	14	7
To return	51	53	28
No definite follow-up	6	21	60
Total	100%	100%	100%
	n=125	n=689	n=1935

NOTES: The practitioner determined seriousness. Cases were eliminated if they provided incomplete data on the particular questions used for this table.

Table 18

PCNPs' use of Physician Consultation, Referral, and Return Appointments, by Seriousness of Condition Treated

Extent of MD Involvement	Very Serious/ Serious	Slightly Serious	Not Serious
First Visits			
Consultation	22%	14%	4%
Referred to MD	13	11	7
To return	35	35	19
No definite follow-up	30	40	69
Total	100%	100%	100%
	n=23	n=186	n=853
Return Visits			
Consultation	11%	10%	8%
Referred to MD	11	11	7
To return	42	46	25
No definite follow-up	37	34	60
Total	100%	100%	100%
	n=38	n=167	n=463

NOTES: The practitioner determined seriousness. Cases were eliminated if they provided incomplete data on the particular questions used for this table.

a signature only. Further, the civilian clinic where the PAs were practicing was constrained by law to structure physician involvement in the treatment process. Chart reviews showed that PAs erred on the side of consultation when not needed, rather than erring by failing to consult when needed. Considering these differences, the 5 to 8 percent consultation rate at the demonstration bases compares closely with this civilian experience.

Another indication of PEs' performance and appropriateness of supervision comes from hospital commanders and supervising physicians at the demonstration bases. During the course of the demonstration, no incidents were reported that indicated that PEs were failing to consult when necessary with their supervisors. Even as the extenders at these bases assumed a significant portion of primary care delivery, base commanders and supervisors who were interviewed expressed confidence in the extenders' ability to judge when a case called for the physician's involvement.

The evidence we have cited suggests that the extenders' care in the demonstration project was adequately supervised.[2] It is unlikely that a more detailed study of supervision behavior would change this conclusion. Consultation is an established part of the extenders' pattern of practice. Extenders' consultations occupy a modest amount of physician time. The amount of consultation does not differ greatly from the amount experienced in a similar civilian setting.

---

[2]The quality of care information presented above provides indirect evidence in support of this conclusion.

VI. CONCLUSION

This report has presented data aimed at evaluating the quality of care delivered by physician's extenders (physician's assistants and primary care nurse practitioners) who functioned in the primary medical clinics of Air Force Hospitals. We have examined the performance of these extenders in relation to the performance of physicians working in the same settings.

We adduced several strands of evidence in evaluating the medical appropriateness of having physician's extenders (PEs) assume a substantial portion of the workload formerly carried principally by physicians. We presented the results of condition-specific, quality-of-care criteria applied to the care provided; reviewed differences in the pattern of ordering tests, procedures, and return visits; and analyzed the supervisory-consultative relationships between physicians and extenders. All conclusions are based on data gathered by encounter form for one month, in the context of an ongoing demonstration project at four Air Force Hospitals.

For the condition-specific, quality-of-care criteria, we looked at five different classes: desirable diagnostic actions, desirable therapeutic actions, undesirable diagnostic actions, undesirable therapeutic actions, and desirable disposition actions. The summary comparison showed that, for a total of 25 out of 28 criteria, PAs' performance statistically equalled or exceeded the physicians' performance; and PCNPs did the same for a total of 14 out of 19 criteria. For most criteria no significant differences were found among the three groups of providers. Also noteworthy was the lack of a systematic deficiency of any group of providers. We conclude that, insofar as we can determine with our criteria, for the types of conditions that they are treating in this setting, PAs and PCNPs do measure up to the performance level of physicians working in the same setting.

Out of 62 criteria, significant differences in performance between PAs and PCNPs occurred for only three of the criteria. We therefore conclude that there is no overall difference in performance.

Furthermore, when 1977 results were compared with 1974 results, there was no evidence of any worsening in PAs' performance—despite classes' having been trained at different times, a number of PAs having been out of school for a longer time and therefore more experienced but potentially less rigorous, and a higher ratio of PAs to supervising physicians in 1977. In addition, based on the 1977 data, we found that recently graduated PAs performed equally as well as earlier graduates. Therefore, the "product" being graduated from the Air Force's in-house PA training program appears to be both consistent and stable.

The condition-specific criteria included a number of measures of quality that we used to detect overuse of certain diagnostic and therapeutic actions. The "utilization analysis" section examined that possibility by comparing utilization rates for physicians, PAs, and PCNPs across a number of tests and procedures. Although observed differences in ordering rates did occur among the three groups of providers, there was no consistent evidence that extenders significantly overburdened the Air Force's care delivery system. Further, with rare exceptions, it was difficult to decide whether observed differences in ordering rates represented neglect or caution, prudence or waste, higher quality or lower quality.

Likewise, we showed that PEs are not generating an inordinate number of return visits, referrals, or hospital admissions.

Finally, in observing the use of the supervisory-consultative channel established between extenders and physicians, we found that the extenders involved the physicians more frequently in complex problems, and with patients whose condition was believed to be serious or very serious. The absolute rate of consultation between extenders and physicians was consistent

with observations in one study made in the civilian sector. Consultation took up only a modest amount of physicians' time. We conclude that the evidence available to us suggests that extenders' care in the demonstration project was adequately supervised.

In view of all these encouraging findings, we believe that the Air Force can deliver the same quality of medical care when PEs (PAs and PCNPs) treat a sizable proportion of the patients formerly treated by physicians, and that no quality bar exists to the continued training and employment of PAs and PCNPs in Air Force outpatient clinics. We find the quality of care they deliver to be acceptable when they are providing care for the types of problems they have been trained to treat. In addition, the PAs' performance, as we have measured it, constitutes a strong endorsement of the Air Force's in-house PA training program.

Appendix A

PATIENT CONTACT RECORD: SAMPLE FORM, AND FREQUENCIES OF CONDITIONS MARKED BY SELECTED PRACTITIONERS, 1977

№ 307603

CARD 1 (1-7)

ARRIVED.	EXAMINED.
(TIME OF PATIENT ARRIVAL) (8-16)	(TIME PATIENT IS CALLED TO EXAMINATION OR TREATMENT ROOM) (17-25)

(26-27) **OUTPATIENT UNIT:**

- ☐ 1. AIR FORCE CLINIC (FAMILY PRACTICE OR GENERAL THERAPY)
- ☐ 2. FLIGHT SURGEON'S CLINIC
- ☐ 3. PHYSICAL EXAM SECTION
- ☐ 4. EMERGENCY ROOM/WALK IN CLINIC
- ☐ 5. SICK CALL
- ☐ 6. ALLERGY
- ☐ 7. DERMATOLOGY
- ☐ 8. ENT
- ☐ 9. EYE CLINIC
- ☐ 10. MEDICINE
- ☐ 11. OB-GYN
- ☐ 12. ORTHOPEDICS
- ☐ 13. PODIATRY
- ☐ 14. PEDIATRICS
- ☐ 15. PSYCHIATRY
- ☐ 16. SURGERY
- ☐ 17. UROLOGY
- ☐ 18. PHYSICAL THERAPY
- ☐ 19. INHALATION THERAPY
- ☐ 20. OTHER: _____

PATIENT CONTACT RECORD

APPROVED: DBMS

PART I Patient Information

TO BE FILLED OUT BY PATIENT

1 (SPONSOR'S) SOCIAL SECURITY NUMBER: (LAST 4 DIGITS)

(33-36) _____

2 SERVICE OF PATIENT OR PATIENT'S SPONSOR:

(37)
- ☐ 1. AIR FORCE
- ☐ 2. ARMY
- ☐ 3. NAVY
- ☐ 4. MARINE CORPS
- ☐ 5. COAST GUARD
- ☐ 6. CADET/APPLICANT FOR MILITARY SERVICE
- ☐ 7. CIVILIAN EMPLOYEE
- ☐ 8. OTHER _____

3 RANK OF PATIENT OR PATIENT'S SPONSOR

(38)
- ☐ 1. ENLISTED
- ☐ 2. OFFICER

4 PATIENT'S DATE OF BIRTH: ____/____/____ (Month / Day / Year)

(39-44)

5 PATIENT'S SEX:

(45)
- ☐ 1. MALE
- ☐ 2. FEMALE

6 PATIENT'S MARITAL STATUS:

(46)
- ☐ 1. NOT APPLICABLE (PATIENT A CHILD)
- ☐ 2. SINGLE
- ☐ 3. MARRIED
- ☐ 4. SEPARATED/DIVORCED
- ☐ 5. WIDOW/WIDOWER

7 PATIENT'S MILITARY STATUS:

(47)
- ☐ 1. SPOUSE OR DEPENDENT OF ACTIVE MILITARY
- ☐ 2. SPOUSE OR DEPENDENT OF RETIRED OR DECEASED MILITARY
- ☐ 3. ACTIVE MILITARY
- ☐ 4. RETIRED MILITARY
- ☐ 5. CIVILIAN EMPLOYEE

8 PATIENT'S FLYING STATUS (ACTIVE DUTY PERSONNEL ONLY):

(48)
- ☐ 1. ON FLYING STATUS
- ☐ 2. SUSPENDED FROM FLYING STATUS
- ☐ 3. NOT ON FLYING STATUS

9 DO YOU (THE PATIENT) LIVE ON THIS BASE?

(49)
- ☐ 1. YES
- ☐ 2. NO

10 IF "NO"

(50-52) APPROXIMATELY HOW MANY MILES AWAY? _____ MILES

11 DID YOU MAKE AN APPOINTMENT FOR THIS VISIT?

(53)
- ☐ 1. YES
- ☐ 2. NO

12 IF YOU HAD AN APPOINTMENT:

(54-55) a. ABOUT HOW MANY DAYS AGO WAS THE APPOINTMENT MADE? (ENTER 0 IF MADE TODAY)
____ DAYS

(56-59) b. WHAT WAS THE TIME OF THE APPOINTMENT? _____ : _____

13 WHAT IS THE MAJOR REASON OR SYMPTOM FOR THIS VISIT?

(60-71)

HEAD AND CHEST AREA:
- ☐ 1. COLD OR RUNNY NOSE
- ☐ 2. COUGH
- ☐ 3. SORE THROAT
- ☐ 4. EARACHE OR EAR DRAINAGE
- ☐ 5. HEADACHE
- ☐ 6. DIZZINESS OR LIGHTHEADEDNESS
- ☐ 7. EYE PROBLEM, OR CAN'T SEE WELL
- ☐ 8. HEART PROBLEM
- ☐ 9. CHEST PAIN

STOMACH (ABDOMINAL) AREA:
- ☐ 10. INDIGESTION, HEARTBURN, DISCOMFORT AFTER EATING
- ☐ 11. STOMACH (ABDOMINAL) PAIN
- ☐ 12. URINARY PROBLEM OR INFECTION
- ☐ 13. PROBLEM WITH SEX AREA OF BODY

BONE, MUSCLE, SKIN:
- ☐ 14. HURT A BONE, JOINT OR MUSCLE
- ☐ 15. BACK ACHE OR NECK PAIN
- ☐ 16. JOINT PAIN (OTHER THAN INJURY) OR ARTHRITIS
- ☐ 17. SKIN CUT, SCRATCH OR BRUISE
- ☐ 18. RASH

GENERAL:
- ☐ 19. FLU
- ☐ 20. FEVER
- ☐ 21. FATIGUE
- ☐ 22. WEAKNESS
- ☐ 23. NERVOUS TENSION, ANXIETY, DEPRESSION
- ☐ 24. OVERWEIGHT (OR WEIGHT GAIN)
- ☐ 25. DIABETES
- ☐ 26. HIGH BLOOD PRESSURE (HYPERTENSION)
- ☐ 27. THYROID PROBLEM
- ☐ 28. CHEMOTHERAPY
- ☐ 29. TO REFILL A PRESCRIPTION
- ☐ 30. TO GET LAB TEST
- ☐ 31. REGULAR (ROUTINE) PHYSICAL EXAM

(72-79) SOMETHING ELSE (WRITE IN): _____

PART II Diagnostic Information

TO BE COMPLETED BY THE HEALTH CARE PROFESSIONAL(S) ONLY

19 a. STATUS OF PATIENT (CHECK ALL BOXES THAT APPLY)

CARD 2 (8) ☐ 1 NEW PATIENT TO CLINIC
(9) ☐ 2 NEW PATIENT TO PRACTITIONER

b. STATUS OF VISIT

(10) ☐ 1 FIRST VISIT (TO ANYONE) FOR MAIN PROBLEM OR PROCEDURE
☐ 2 RETURN VISIT (TO ANYONE) FOR MAIN PROBLEM OR PROCEDURE
☐ 3 PRESCRIPTION REFILL ONLY

20 ANY PROPHYLACTIC PROCEDURES PERFORMED? IF YES, CHECK APPLICABLE BOX

☐ 1 PERIODIC ADULT PHYSICAL EXAM
(11 12) ☐ 2 PERIODIC WELL CHILD EXAM
☐ 3 FLIGHT PHYSICAL (CLASS 1)
☐ 4 FLIGHT PHYSICAL (CLASS 2 OR 3)
☐ 5 SEPARATION/RETIREMENT PHYSICAL
☐ 6 OTHER NON FLYING (ADMINISTRATIVE) PHYSICAL
☐ 7 ROUTINE EYE EXAM
☐ 8 ROUTINE GYN EXAM
☐ 9 PRENATAL VISIT
☐ 10 POSTPARTUM VISIT
☐ 11 PRE OP VISIT
☐ 12 POST OP FOLLOW UP
☐ 13 WEIGHT CHECK ONLY
☐ 14 HAZARDOUS NOISE EXAM
☐ 15 OTHER _____

21 ID NUMBER OF PERSON(S) TREATING PATIENT (ENTER IN APPROPRIATE SPACE)

(13 27) _____ / _____ / _____ / _____ / _____
DR PA NURSE CORPSMAN OTHER

22 PROBLEMS/AREAS TREATED (CHECK ALL BOXES THAT APPLY)

1. COMMUNICABLE DISEASES (SEE ALSO SECTIONS 8, 10, 11, 12)

(28 42) ☐ 5 INFECTIOUS INTESTINAL (INCLUDES INFECTIOUS DIARRHEA)
☐ 23 VIRAL SYNDROME WITH GASTROENTERITIS
☐ 245 VIRAL SYNDROME WITHOUT GASTROENTERITIS
☐ 11 13, 15 MEASLES, MUMPS, CHICKEN POX
☐ 16 HEPATITIS OR EXPOSURE TO HEPATITIS
☐ 17 INFECTIOUS MONONUCLEOSIS
☐ 4 GONORRHEA (OR EXPOSURE TO GONORRHEA)
☐ 901 OTHER _____

2. NEOPLASMS

☐ 50 68 MALIGNANT NEOPLASM
☐ 831 BENIGN SKIN (INCLUDES MOLE/NEVUS)
☐ 70 74 BENIGN (OTHER THAN SKIN)
☐ 902 OTHER _____

3. ALLERGIC, ENDOCRINE, METABOLIC, NUTRITIONAL

☐ 85, 95, 257 HAY FEVER/ALLERGIES
☐ 88 ASTHMA
☐ 88 90 THYROID DISEASE
☐ 91 DIABETES MELLITUS
☐ 101 OBESITY
☐ 903 OTHER _____

4. BLOOD AND IMMUNOLOGICAL ORGANS

☐ 111 IRON DEFICIENCY (HYPOCHROMIC) ANEMIA
☐ 110, 112, 122 OTHER ANEMIAS
☐ 904 OTHER _____

5. NERVOUS SYSTEM, MENTAL ILLNESS, PERSONALITY DISORDERS

☐ 124 129 PSYCHOSIS
☐ 130, 134 ANXIETY OR DEPRESSIVE NEUROSIS
☐ 633 SEXUAL DYSFUNCTION
600 ADULT SITUATION DISTURBANCE (E.G., MARITAL)
☐ 139 DRUG DEPENDENCE/ABUSE
☐ 141 ALCOHOL ABUSE OR ALCOHOLISM
☐ 147 TENSION HEADACHE
☐ 159 MIGRAINE HEADACHE (OR OTHER MIGRAINE MANIFESTATIONS)
☐ 454 OTHER HEADACHE
☐ 155 VASCULAR LESIONS (INCLUDES STROKE, CEREBRAL ARTERIOSCLEROSIS)
☐ 158, 197 EPILEPSY, CONVULSIONS
☐ 156 157, 159 160, 165 169 OTHER NERVOUS SYSTEM DISEASES
150 PROBLEM OF DEVELOPMENT RETARDATION OR BEHAVIOR
☐ 905 OTHER _____

6. EYE AND EAR

170 CONJUNCTIVITIS OR OPHTHALMIA
☐ 176 REFRACTIVE ERRORS
☐ 178, 605 STRABISMUS, TROPIA OR PHORIA
☐ 171 175, 177, 179 181 OTHER EYE DISEASES
☐ 182 OTITIS EXTERNA
☐ 183, 184 OTITIS MEDIA (EXCLUDES SEROUS)
☐ 648 SEROUS OTITIS MEDIA
☐ 187 WAX IN EAR
☐ 761, 185 186, 188 190 OTHER DISEASES OF THE EAR
☐ 906 OTHER _____

7. CARDIOVASCULAR

☐ 211, 212 215 ISCHEMIC HEART DISEASES (INCLUDING ANGINA PECTORIS, ASHD)
☐ 214 ARRHYTHMIAS OR HEART BLOCK
☐ 634 HEART MURMUR
☐ 213, 215, 217 OTHER HEART DISEASES
☐ 218 HYPERTENSION (HBP)
☐ 233 SYNCOPE
☐ 225 HEMORRHOIDS
☐ 224 VARICOSE VEINS
☐ 907 OTHER _____

8. RESPIRATORY

☐ 240 CORYZA (NON FEBRILE COMMON COLD)
☐ 241 FEBRILE COLD
☐ 245 INFLUENZA, RESPIRATORY FLU SYNDROME
☐ 242 SORE THROAT (PHARYNGITIS OR TONSILLITIS)
☐ 243 ACUTE SINUSITIS
☐ 250 CHRONIC SINUSITIS
86 ASTHMA
247 258 ACUTE BRONCHITIS (OR BRONCHIOLITIS)
☐ 246 PNEUMONIA, PNEUMONITIS
248, 255 CHRONIC BRONCHITIS/EMPHYSEMA/COPD
267 COUGH ONLY
263 NOSE BLEED
908 OTHER _____

9 DIGESTIVE

☐ 273 274 284 294	ESOPHAGITIS GASTRITIS INDIGESTION HIATAL HERNIA
☐ 277 279	ULCER DISEASE (STOMACH OR DUODENUM)
☐ 281	OTHER DISEASES OF ESOPHAGUS STOMACH DUODENUM
☐ 5	INFECTIOUS INTESTINAL (INCLUDES INFECTIOUS DIARRHEA)
☐ 23	VIRAL SYNDROME WITH GASTROENTERITIS
☐ 301, 306	ABDOMINAL PAIN (NOT OTHERWISE SPECIFIED)
☐ 280	FUNCTIONAL UPPER GI DISTRESS
☐ 304	FUNCTIONAL LARGE BOWEL DISTRESS (IRRITABLE SPASTIC COLON)
☐ 286 287	CHOLELITHIASIS, CHOLECYSTITIS
☐ 282, 285 309	OTHER DISEASES OF INTESTINE AND PERITONEUM
☐ 303	DIARRHEA
☐ 225	HEMORRHOIDS
☐ 283	HERNIA (INGUINAL, FEMORAL, UMBILICAL
☐ 909	OTHER _____

10. GENITO URINARY SYSTEM (SEE ALSO SECTION 1)

☐ 313, 314	URINARY TRACT INFECTION (CYSTITIS PYELITIS/PYELONEPHRITIS)
☐ 315	NONSPECIFIC URETHRITIS (NON GONOCOCCAL)
☐ 4	GONOCOCCAL URETHRITIS, CERVICITIS, OR SALPINGITIS
☐ 317	OTHER DISEASES OF URINARY SYSTEM
☐ 316, 318	PROSTATITIS OR BENIGN PROSTATIC HYPERTROPHY
☐ 319 321, 331	OTHER DISEASES OF REPRODUCTIVE SYSTEM
☐ 335	VULVITIS, VAGINITIS AND CERVICITIS (NON VENEREAL)
☐ 327 330, 334	DISORDERS OF MENSTRUATION DYSFUNCTIONAL UTERINE BLEEDING
☐ 329	MENOPAUSAL SYMPTOMS
☐ 585, 586	FAMILY PLANNING CONTRACEPTION/INFERTILITY
☐ 332	CERVICAL EROSION
☐ 322	BREAST MASS OR BREAST DISEASE (EXCLUDING MALIGNANCY)
☐ 910	OTHER _____

12. SKIN AND SUPERFICIAL TISSUE (SEE ALSO SECTION 16)

☐ 21	FUNGAL SKIN INFECTION, DERMATOPHYTOSIS
☐ 371, 372, 374	CELLULITIS (INCLUDING LYMPHANGITIS)
☐ 375	IMPETIGO
☐ 369	ACNE
☐ 370, 373, 377	OTHER LOCAL INFECTIONS OF SKIN AND SUBCUTANEOUS TISSUE
☐ 378-381	DERMATITIS (INCLUDING ECZEMA)
☐ 368	PITYRIASIS ROSEA
☐ 644	DRUG RASH
☐ 398	RASH (OTHERWISE UNSPECIFIED)
☐ 631	MOLE, NEVUS
☐ 25	WARTS
☐ 384, 385	CORNS/OTHER HYPERTROPHIC/ATROPHIC SKIN CONDITIONS
☐ 387	DISEASES OF NAIL AND NAIL BED (EXCLUDING FUNGUS)
☐ 389	DISEASES OF SWEAT AND SEBACEOUS GLANDS (INCLUDING SEBACEOUS CYST)
☐ 912	OTHER _____

13. BONES, JOINTS, MUSCLES

☐ 406	RHEUMATOID ARTHRITIS
☐ 405	OSTEOARTHRITIS
☐ 407-409	OTHER ARTHRITIS/RHEUMATISM, INCLUDING POST TRAUMA
☐ 420-422	BURSITIS, TENOSYNOVITIS, SYNOVITIS
☐ 424-425	BACKACHE ALONE, BACK PAIN ALONE
☐ 423	BACKACHE WITH SCIATICA
☐ 428	PAIN IN JOINT (ARTHRALGIA)
☐ 479	DISLOCATION, UPPER EXTREMITY
☐ 640	DISLOCATION, LOWER EXTREMITY
☐ 642	TRAUMA TO HEAD
☐ 473-477	FRACTURE OF UPPER LIMB
☐ 488, 489	FRACTURE OF LOWER LIMB
☐ 495	OTHER FRACTURE
☐ 807	SPRAIN/STRAIN UPPER LIMB
☐ 808	SPRAIN/STRAIN LOWER LIMB
☐ 809	SPRAIN/SYRAIN NECK/BACK
☐ 643	MUSCLE PAIN, MUSCLE CRAMPS
☐ 913	OTHER _____

16 ACCIDENTS, POISONINGS AND VIOLENCE (SEE ALSO SECTION 13)

465 484	FOREIGN BODY (DEFINITE OR POSSIBLE)
483	LACERATIONS CONTUSIONS ABRASIONS SUPERFICIAL INJURIES
394	INSECT BITES
611	ANIMAL BITES
485 487	BURNS
488 491	POISONING OVERDOSE
917	OTHER _____

17 SIGNS SYMPTOMS AND ILL DEFINED CONDITIONS

455	MALINGERING
458	FEVER OF UNKNOWN ORIGIN
646	PAIN
231 271	CHEST PAIN (UNKNOWN ETIOLOGY)
647	NO PATHOLOGY AT THIS TIME (WELL PATIENT)
916	NO DEFINITE DIAGNOSIS AT THIS TIME

19 OTHER

919	OTHER _____

23 SERIOUSNESS OF MAIN CONDITION TREATED (CHECK ONE)

(43) ☐ 1 VERY SERIOUS
☐ 2 SERIOUS
☐ 3 SLIGHTLY SERIOUS
☐ 4 NOT SERIOUS

24 *DISPOSITION OF VISIT*

(44 49) ☐ 1 NO FORMAL FOLLOW UP PLANNED
☐ 2 RETURN PRN (POSSIBLE FOLLOW UP)
☐ 3 DEFINITE TELEPHONE FOLLOW UP
☐ 4 DEFINITE RETURN APPOINTMENT
☐ 5 REFER TO OTHER PROFESSIONAL OR CLINIC ON THIS VISIT (SPECIFY BY ID NUMBER OR NAME)
_____
☐ 6 REFER TO OTHER PROFESSIONAL OR CLINIC FOR FUTURE APPOINTMENT (SPECIFY BY ID NUMBER OR NAME)
_____
☐ 7 *REFER TO CHAMPUS*
☐ 8 ADMIT TO QUARTERS
☐ 9 ADMIT TO HOSPITAL
☐ 10 RETURN TO REFERRING CLINIC (SPECIFY BY NAME)
_____
☐ 11 OTHER _____

(50 52) _____

25 a. DID YOU SPEAK OR CONSULT WITH ANOTHER HEALTH CARE PROFESSIONAL CONCERNING THIS PATIENT ?

(53) ☐ 1 YES – AND OTHER PROFESSIONAL SAW PATIENT ON THIS VISIT
☐ 2 YES – SPOKE, BUT OTHER PROFESSIONAL DID NOT SEE PATIENT ON THIS VISIT
☐ 3 YES – BUT ONLY TO HAVE A PRESCRIPTION OR ORDER COUNTERSIGNED
☐ 4 NO

(54 56) **b. IF YES, ENTER NAME OR ID NUMBER OF OTHER PROFESSIONAL CONSULTED:** _____
_____

PART III Treatment Information

IN THE QUESTIONS WHICH FOLLOW, EACH HEALTH PROFESSIONAL SHOULD CHECK OFF APPROPRIATE BOXES IN HIS OR HER COLUMN

26 TESTS (CHECK ALL APPLICABLE BOXES)

	a. DIAGNOSTIC TESTS (ORDERED OR PERFORMED)	DR 1	PA 2	NURSE 3	CORPS MAN 4	OTHER 5
(57)	AUDIOGRAM	☐	☐	☐	☐	☐
(58)	CULTURE - GENITAL TRACT, OR WET PREP SMEAR, OR GRAM STAIN	☐	☐	☐	☐	☐
(59)	CULTURE - THROAT	☐	☐	☐	☐	☐
(60)	CULTURE - URINE	☐	☐	☐	☐	☐
(61)	CULTURE - OTHER SOURCES (SPUTUM, STOOL, ETC.)	☐	☐	☐	☐	☐
(62)	EXERCISE (STRESS) TEST	☐	☐	☐	☐	☐
(63)	PAP SMEAR	☐	☐	☐	☐	☐
(64)	PULMONARY FUNCTION TESTS	☐	☐	☐	☐	☐
(65)	SKIN TESTS (ALLERGY, TB, ETC.)	☐	☐	☐	☐	☐
(67)	URINALYSIS	☐	☐	☐	☐	☐
(68)	OTHER _____	☐	☐	☐	☐	☐
CARD 3	b. LAB TESTS ORDERED					
(8)	HEMOGLOBIN/HEMATOCRIT (ONLY)	☐	☐	☐	☐	☐
(9)	WBC/WHITE BLOOD COUNT (WITH OR WITHOUT DIFF.)	☐	☐	☐	☐	☐
(10)	COMPLETE BLOOD COUNT/CBC	☐	☐	☐	☐	☐
(11)	SYPHILIS SEROLOGY, VDRL	☐	☐	☐	☐	☐
(12)	MONOSPOT OR HETEROPHILE	☐	☐	☐	☐	☐
(13)	URINALYSIS	☐	☐	☐	☐	☐
(14)	BLOOD SUGAR (SINGLE OR TOLERANCE TEST)	☐	☐	☐	☐	☐
(15)	CHOLESTEROL/TRIGLYCERIDES (LIPIDS)	☐	☐	☐	☐	☐
(16)	BUN/CREATININE	☐	☐	☐	☐	☐
(17)	ELECTROLYTES (ONE OR MORE)	☐	☐	☐	☐	☐
(18)	LIVER FUNCTION TEST(S)	☐	☐	☐	☐	☐
(19)	THYROID (T3, T4, TSH, ETC.)	☐	☐	☐	☐	☐
(20)	EKG	☐	☐	☐	☐	☐
(21)	X RAY - CHEST	☐	☐	☐	☐	☐
(22)	X RAY OTHER THAN CHEST	☐	☐	☐	☐	☐
(23)	OTHER _____	☐	☐	☐	☐	☐
	c. SELECTED EXAMINATION PROCEDURES PERFORMED					
(24)	BLOOD PRESSURE MEASURED	☐	☐	☐	☐	☐
(25)	LUNGS AUSCULATED	☐	☐	☐	☐	☐
(26)	HEART AUSCULATED	☐	☐	☐	☐	☐
(27)	ABDOMEN PALPATED	☐	☐	☐	☐	☐
(28)	PELVIC EXAMINATION DONE	☐	☐	☐	☐	☐

27 TREATMENT (CHECK ALL APPLICABLE ITEMS):

	a. PHYSICAL MEASURES:	DR 1	PA 2	NURSE 3	CORPS MAN 4	OTHER 5
(29)	SUTURE PLACEMENT OR REMOVAL	☐	☐	☐	☐	☐
(30)	APPLY, CHECK, CHANGE DRESSING	☐	☐	☐	☐	☐
(31)	REMOVE FOREIGN BODY	☐	☐	☐	☐	☐
(32)	INCISION, EXCISION OR BIOPSY	☐	☐	☐	☐	☐
(33)	ORTHOPEDIC PROCEDURE	☐	☐	☐	☐	☐
(34)	ORDER/GIVE PHYSICAL THERAPY	☐	☐	☐	☐	☐
(35)	IMMUNIZATION OR VACCINATION (ORDERED OR DONE)	☐	☐	☐	☐	☐
(36)	INJECTION OTHER THAN IMMUN / VACC (ORDERED OR DONE)	☐	☐	☐	☐	☐
(37)	CONTRACEPTIVE PROCEDURE (INCLUDES IUD)	☐	☐	☐	☐	☐
(38)	OTHER _____	☐	☐	☐	☐	☐
	b. SELECTED MEDICATIONS (WHETHER STARTED OR CONTINUED):					
(39)	ANALGESIC - NARCOTIC (OR COMBINATION NON NARCOTIC WITH NARCOTIC)	☐	☐	☐	☐	☐
(40)	ANALGESIC - NON NARCOTIC	☐	☐	☐	☐	☐
(41)	ANTACID, ANTI EMETIC, ANTIDIARRHEAL	☐	☐	☐	☐	☐
(42)	ANTI ANEMIA (HEMATINIC)					
	ANTIBIOTICS					
(43)	ANY PENICILLIN (INCLUDING AMPICILLIN) OR ERYTHROMYCIN	☐	☐	☐	☐	☐
(44)	ANY TETRACYCLINE	☐	☐	☐	☐	☐
(45)	ANY SULFA DRUG	☐	☐	☐	☐	☐
(46)	ANY OTHER ANTIBIOTIC	☐	☐	☐	☐	☐
(47)	ANTIHISTAMINIC/DECONGESTANT/ ANTITUSSIVE/ANTI COLD	☐	☐	☐	☐	☐
(48)	ANTIHYPERTENSIVE AND/OR DIURETIC	☐	☐	☐	☐	☐
(49)	ANTI INFLAMMATORY	☐	☐	☐	☐	☐
(50)	BRONCHIAL DILATOR	☐	☐	☐	☐	☐
(51)	CARDIOVASCULAR (DIG., ANTI ANGINAL, ANTI ARRHYTHMIC)	☐	☐	☐	☐	☐
(52)	CONTRACEPTIVE (ORAL) OR ESTROGEN PREPARATION	☐	☐	☐	☐	☐
(53)	INSULIN OR ORAL HYPOGLYCEMIC AGENT	☐	☐	☐	☐	☐
(54)	SEDATIVE/SLEEP MED	☐	☐	☐	☐	☐
(55)	STEROID - ORAL OR INJECTED	☐	☐	☐	☐	☐
(56)	STEROID - TOPICAL OR INHALED	☐	☐	☐	☐	☐
(57)	THYROID RELATED MED	☐	☐	☐	☐	☐
(58)	TRANQUILIZER OR ANTI DEPRESSANT	☐	☐	☐	☐	☐
(59)	VITAMINS AND MINERALS INC. IRON	☐	☐	☐	☐	☐
(60)	OTHER _____	☐	☐	☐	☐	☐
	c. COUNSELING AND OTHER					
(61)	COUNSEL ABOUT DISEASE, MEDICATION, OR OTHER TREATMENT	☐	☐	☐	☐	☐
(62)	ADVICE ABOUT CONTRACEPTION	☐	☐	☐	☐	☐
(63)	PSYCHOLOGICAL COUNSELING	☐	☐	☐	☐	☐
(64)	CHAPERONING	☐	☐	☐	☐	☐
(65)	ADMINISTRATIVE ACTION	☐	☐	☐	☐	☐
(66)	OTHER _____	☐	☐	☐	☐	☐

PART IV Time Information

PATIENT SHOULD ANSWER THE FOLLOWING QUESTIONS AT THE END OF THE VISIT

14 DURING YOUR VISIT, APPROXIMATELY HOW MUCH TIME WAS SPENT WITH THE DOCTOR, IF ANY?

(67) NONE 0 ☐ | LESS THAN A MIN 1 ☐ | 1-5 MIN 2 ☐ | 5-10 MIN 3 ☐ | 10-20 MIN 4 ☐ | 20-30 MIN 5 ☐ | MORE THAN 30 MIN 6 ☐

15 HOW MUCH TIME WAS SPENT WITH THE PHYSICIAN ASSISTANT (IDENTIFIED BY THE BLUE EMBLEM "PA" ON HIS JACKET), IF ANY?

(68) NONE 0 ☐ | LESS THAN A MIN 1 ☐ | 1-5 MIN 2 ☐ | 5-10 MIN 3 ☐ | 10-20 MIN 4 ☐ | 20-30 MIN 5 ☐ | MORE THAN 30 MIN 6 ☐

16 HOW MUCH TIME WAS SPENT WITH THE NURSE, OR NURSE PRACTITIONER, IF ANY?

(69) NONE 0 ☐ | LESS THAN A MIN 1 ☐ | 1-5 MIN 2 ☐ | 5-10 MIN 3 ☐ | 10-20 MIN 4 ☐ | 20-30 MIN 5 ☐ | MORE THAN 30 MIN 6 ☐

17 HOW MUCH TIME WAS SPENT WITH THE CORPSMAN, IF ANY?

(70) NONE 0 ☐ | LESS THAN A MIN 1 ☐ | 1-5 MIN 2 ☐ | 5-10 MIN 3 ☐ | 10-20 MIN 4 ☐ | 20-30 MIN 5 ☐ | MORE THAN 30 MIN 6 ☐

18 HOW MUCH TIME WAS SPENT WITH OTHER HEALTH CARE PROFESSIONALS, IF ANY?

(71) NONE 0 ☐ | LESS THAN A MIN 1 ☐ | 1-5 MIN 2 ☐ | 5-10 MIN 3 ☐ | 10-20 MIN 4 ☐ | 20-30 MIN 5 ☐ | MORE THAN 30 MIN 6 ☐

Table A.1

Frequencies of Conditions Marked by Selected Practitioners at Dyess, Fairchild, Chanute, and Nellis Air Force Bases: 1977 Patient Contact Record

Most Serious Diagnosis: Code Number and Description		MD Only	PA Only	PCNP Only
1. Communicable Diseases				
*5	Infectious intestinal (incl. infectious diarrhea)	4	5	1
*23	Viral syndrome w/gastroenteritis	104	160	51
*245	Viral syndrome w/o gastroenteritis	148	191	52
11-13, 15	Measles, mumps, chicken pox	8	11	7
16	Hepatitis or exposure to hepatitis	11	8	4
17	Infectious mononucleosis	1	15	2
*4	Gonorrhea (or exposure to gonorrhea)	17	28	10
901	Other	21	18	1
	Total	314	436	128
	% of practitioner total	6.0%	5.3%	6.2%
2. Neoplasms				
50-68	Malignant neoplasm	30	4	2
*631	Benign skin (incl. mole/nevus)	8	24	11
70-74	Benign (other than skin)	3	4	5
902	Other	7	5	3
	Total	48	37	21
	% of practitioner total	0.9%	0.5%	1.0%
3. Allergic Endocrine, Metabolic, Nutritional				
85, 95, 257	Hay fever/allergies	100	248	61
*86	Asthma	59	61	14
88-90	Thyroid disease	57	93	15
91	Diabetes mellitus	90	78	10
101	Obesity	84	65	9
903	Other	41	39	13
	Total	431	584	122
	% of practitioner total	8.2%	7.1%	5.9%
4. Blood and Immunological Organs				
111	Iron-deficiency (hypochromic) anemia	14	12	2
110, 112, 122	Other anemias	4	6	1
904	Other	14	8	2
	Total	32	26	5
	% of practitioner total	0.6%	0.3%	0.2%

Table A.1—continued

Most Serious Diagnosis: Code Number and Description		MD Only	PA Only	PCNP Only
5. *Nervous System, Mental Illness, Personality Disorders*				
124-129	Psychosis	5	2	0
130, 134	Anxiety or depressive neurosis	50	38	10
633	Sexual dysfunction	1	1	0
600	Adult situation disturbance (e.g., marital)	1	22	3
139	Drug dependence/abuse	6	2	0
141	Alcohol abuse or alcoholism	12	3	0
147	Tension headache	19	39	6
159	Migraine headache (or other migraine manifestations)	25	29	8
454	Other headache	18	44	11
155	Vascular lesions (incl. stroke, cerebral arteriosclerosis)	10	3	0
158, 197	Epilepsy convulsions	30	17	2
156-157, 159-160, 165-169	*Other nervous system diseases*	29	24	15
150	Problem of development, retardation, or behavior	1	4	1
905	Other	17	24	3
	Total	224	252	59
	% of practitioner total	4.3%	3.1%	2.9%
6. *Eye and Ear*				
170	Conjunctivitis or ophthalmia	50	109	41
176	Refractive errors	4	3	1
178, 605	Strabismus, *tropia, or phoria*	1	3	2
171-175, 177, 179-181	Other eye diseases	52	53	15
182	Otitis externa	35	68	21
183, 184	Otitis media (*excl. serous*)	104	197	50
648	Serous otitis media	25	247	54
187	Wax in ear	15	34	17
161, 185-186, 188-190	Other diseases of the ear	24	60	13
906	Other	29	13	7
	Total	339	787	221
	% of practitioner total	6.5%	9.6%	10.7%
7. *Cardiovascular*				
211, 212-215	Ischemic heart diseases (incl. angina pectoris, ASHD)	116	13	8
214	Arrhythmias or heart block	62	16	6
634	Heart murmur	43	12	4
213, 215, 217	Other heart disease	15	19	1
218	Hypertension (HBP)	220	417	101
233	Syncope	5	6	2
*225	Hemorrhoids	13	45	10
224	Varicose veins	4	8	3
907	Other	24	25	6
	Total	502	561	141
	% of practitioner total	9.6%	6.8%	6.8%

Table A.1—continued

Most Serious Diagnosis: Code Number and Description		MD Only	PA Only	PCNP Only
8. Respiratory				
240	Coryza (nonfebrile common cold)	225	587	176
241	Febrile cold	87	86	65
245	Influenza, respiratory flu syndrome	148	191	52
242	Sore throat (tonsillitis or pharyngitis)	241	429	113
243	Acute sinusitis	46	101	22
250	Chronic sinusitis	31	45	14
86	Asthma	59	61	14
247, 258	Acute bronchitis (or bronchiolitis)	45	189	29
246	Pneumonia, pneumonitis	14	58	23
248, 255	Chronic bronchitis/emphysema/COPD	19	40	10
267	Cough only	18	39	12
263	Nose bleed	7	8	3
908	Other	37	46	8
	Total	977	1880	541
	% of practitioner total	18.7%	22.9%	26.2%
9. Digestive				
273, 274, 284, 294	Esophagitis, gastritis, indigestion, hiatal hernia	40	96	21
277-279	Ulcer disease (stomach or duodenum)	18	36	4
281	Other diseases of esophagus, stomach, duodenum	3	15	3
5	Infectious intestinal (incl. infectious diarrnea)	4	5	1
23	Viral syndrome w/gastroenteritis	104	160	51
301, 306	Abdominal pain (not otherwise specified)	36	82	29
280	Functional upper GI distress	14	18	5
304	Functional large bowel distress (irritable/spastic colon)	7	27	5
286, 287	Cholelithiasis, cholecystitis	10	12	4
282, 285, 309	Other diseases of intestine and peritoneum	22	13	1
303	Diarrhea	18	22	7
225	Hemorrhoids	13	45	10
283	Hernia--inguinal, femoral, umbilical	2	15	3
909	Other	53	52	19
	Total	344	598	163
	% of practitioner total	6.6%	7.3%	7.9%

Table A.1—continued

Most Serious Diagnosis: Code Number and Description		MD Only	PA Only	PCNP Only
10. Genito-Urinary System				
313, 314	Urinary tract infection (*cystitis*/ pyelitis/pyelonephritis)	72	193	54
315	Nonspecific urethritis (nongonococcal)	19	55	14
4	Gonococcal urethritis, cervicitis, or salpingitis	17	28	10
317	Other diseases of urinary system	18	29	9
316, 318	Prostatitis or benign prostatic hypertrophy	7	16	2
319-321, 331	Other diseases of reproductive system	14	33	7
335	*Vulvitis, vaginitis, and cervicitis* (nonvenereal)	8	19	6
327-330, 334	Disorders of menstruation, dysfunctional uterine bleeding	17	12	4
329	Menopausal symptoms	1	3	0
585, 586	Family planning/contraception/ infertility	2	12	1
332	Cervical erosion	0	0	0
322	Breast mass or breast disease (excl. malignancy)	9	24	17
910	Other	27	34	10
	Total	211	458	134
	% of practitioner total	4.0%	5.6%	6.5%
12. Skin and Superficial Tissue				
21	Fungal skin infection, dermatophytosis	44	83	22
371, 372, 374	Cellulitis (incl. lymphangitis)	10	43	3
375	Impetigo	3	19	6
369	Acne	41	137	19
370, 373, 377	Other local infections of skin and subcutaneous tissue	30	92	36
378-381	Dermatitis (incl. eczema)	55	148	33
368	Pityriasis rosea	2	6	1
644	Drug rash	3	4	2
398	Rash (otherwise unspecified)	42	125	20
631	Mole, nevus	8	24	11
25	Warts	21	42	4
384, 385	Corns/other hypertrophic/atrophic skin conditions	8	19	5
387	Diseases of nail and nailbed (excl. fungus)	7	15	6
389	Diseases of sweat and sebaceous glands (incl. sebaceous cyst)	18	53	4
912	Other	37	72	11
	Total	329	882	183
	% of practitioner total	6.3%	10.7%	8.9%

Table A.1—continued

Most Serious Diagnosis: Code Number and Description		MD Only	PA Only	PCNP Only
13. Bones, Joints, Muscles				
406	Rheumatoid arthritis	38	25	1
405	Osteoarthritis	12	74	3
407-409	Other arthritis/rheumatism, incl. post-trauma	27	59	14
420-422	Bursitis, tenosynovitis, synovitis	34	106	22
424-425	Backache alone, back pain alone	64	189	57
423	Backache with sciatica	22	33	12
428	Pain in joint (arthralgia)	29	97	25
479	Dislocation, upper extremity	5	4	0
640	Dislocation, lower extremity	0	1	0
642	Trauma to head	26	17	2
473-477	Fracture of upper limb	26	25	4
468, 469	Fracture of lower limb	15	9	1
495	Other fracture	11	10	1
607	Sprain/strain upper limb	59	64	16
608	Sprain/strain lower limb	81	156	23
609	Sprain/strain neck/back	58	71	22
643	Muscle pain, muscle cramps	59	132	37
913	Other	86	133	40
	Total	652	1205	280
	% of practitioner total	12.5%	14.7%	13.6%
16. Accidents, Poisonings, and Violence				
465, 484	Foreign body (definite or possible)	9	9	0
483	Lacerations, contusions, abrasions, superficial injuries	186	167	22
394	Insect bites	1	3	1
611	Animal bites	17	6	1
485-487	Burns	14	16	4
488-491	Poisoning, overdose	10	0	0
917	Other	18	19	3
	Total	255	220	31
	% of practitioner total	4.9%	2.7%	1.5%
17. Signs, Symptoms, Ill-defined Conditions				
455	Malingering	0	0	0
458	Fever of unknown origin	7	6	1
646	Pain	3	13	5
231, 271	Chest pain (unknown etiology)	28	19	5
647	No pathology at this time (well patient)	15	89	24
916	No definite diagnosis at this time	58	94	50
	Total	111	221	85
	% of practitioner total	2.1%	2.7%	4.1%
19. Other				
919	Other	38	23	15
	% of practitioner total	0.7%	0.3%	0.7%

Table A.1—continued

	Most Serious Diagnosis: Code Number and Description	MD Only	PA Only	PCNP Only
---	*No diagnosis* listed	778	564	81
	% of practitioner total	14.9%	6.9%	3.9%
	Total, all diagnoses	5585	8734	2210
	Minus double-counted diagnoses ...	-353	-514	-149
	Practitioner total	5232	8220	2061

*These diagnoses are repeated under later categories; therefore they are double-counted in this table and cause the practitioner totals to exceed 100 percent.

Appendix B

STATISTICAL TECHNIQUES

TESTS FOR INDIVIDUAL CRITERIA

We used two alternative approaches in testing whether practitioner performance on each individual criterion differed significantly. The first technique compares the compliance rates for all visits (under the criterion) to one practitioner type with the compliance rates for another practitioner type. The second technique uses logit regressions to estimate differences in compliance rates due to practitioner type and also due to the practitioners' base. Unlike the compliance rate techniques, the regression technique allows us to test for differences among practitioners while controlling for base differences.

Compliance Rates

For each criterion we investigated, we calculated rates for each of the three groups of providers: physicians, physician's assistants (PAs), and primary care nurse practitioners (PCNPs). These rates reflect the number of times providers in a group performed the action specified by each quality-of-care criterion, divided by the number of visits to the group's providers that apply to the criterion. Here we describe how we determined whether the observed differences in performance rates are statistically significant, i.e., are likely to reflect real differences among practitioner types.

A group's performance depends on the set of decisions each group member makes. Each practitioner will see a number of patients for a particular problem. The practitioner's pattern of treatment for this problem determines his or her compliance level (p_{ij}) for the criterion applying to the problem; thus, p_{ij} is the underlying per-visit probability that the ith individual of the jth practitioner group will comply. Expected compliance by a group, p_j, depends on the individual compliance of its members and on how the caseload is distributed among the group members:

$$p_j = \sum_{i=1}^{m_j} w_{ij} p_j$$

For visits described by a particular criterion, w_{ij} represents the probability the visit is handled by the ith individual in the jth practitioner group; m_j is the number of individuals in the jth group.

Our test for statistical significance directly compares group performances, the p_j's, two at a time. To use the test, we must assume that p_j represents the probability of compliance for each and every member of group j, ignoring the fact that within group j, members vary in their compliance with a particular criterion. With this assumption, we can define compliance as the result of n_j simple Bernoulli trials with probability of success on each trial p_j, n_j being the number of visits to practitioners from the jth group where the criterion applies. The number of successes, or compliances, in n_j visits is a binomial random variable with parameters p_j and n_j.

We cannot actually observe the underlying compliance probabilities for individuals (p_{ij}) or groups (p_j). Instead we estimate p_j by calculating the actual compliance rates (p_j) achieved during our data collection period.

We can test whether the observed difference between two groups' performances is statistically large by standard tests of the difference in two proportions. The test statistic used is:

$$d_{jk} = \frac{\hat{p}_j - \hat{p}_k}{\sqrt{\frac{n_j + n_k}{n_j n_k} (\hat{\bar{p}}) (1 - \hat{\bar{p}})}}$$

where d_{jk} = test statistic for differences between groups j and k;

$\hat{\bar{p}}$ = combined compliance rate of both groups = $\frac{n_j \hat{p}_j + n_k \hat{p}_k}{n_j + n_k}$

With large enough numbers of visits, d_{jk} is distributed approximately as a standard normal variable.

Selection of the critical value (K) for the test depends on the fact that two comparisons against physicians' performances are made, one for PAs and one for PCNPs. To limit to the 5 percent level the chances of erroneously concluding that both PAs and PCNPs differ, we selected a critical value of 2.24 for each of the comparisons, MD with PA and MD with PCNP. This gives a 5 percent limit to the chances of erring in the pair of conclusions.[1]

Power of the Test

Many of the criteria that are reported in Appendix C have so few cases that the significance tests may have little meaning when they fail to reject the hypothesis of equality. With so few observations, the chances of accepting the hypothesis of equality between practitioner types can be very large even if true differences exist between the practitioners' compliance rates.

Criteria that were used in Table 4 for the summary comparisons were chosen based, in part, on the power of each test. We included in the summary only those criteria with sufficient observations to result in a 75 percent chance of getting a significant difference when the true difference between compliance rates ($p_j - p_k$) was 0.25 or greater. Thus we protect ourselves from the chances of erroneously concluding that the compliance rates are equal when they actually differ by a considerable amount.

To calculate the power of the test, we begin by specifying the alternate hypothesis, H_a: $p_j \neq p_k$. The power of the test (ϕ) is the probability of rejecting the null hypothesis (H_o: $p_j = p_k$), given the true values of p_j and p_k and the sizes of the two samples n_j and n_k. Thus, ϕ is the probability the test statistic, d_{jk}, previously calculated, exceeds the selected critical value K, 2.24 in our case.

We want to calculate

$$\phi = \Pr(d_{jk} > K).$$

Multiplying both sides of this expression by the denominator of d_{jk} and substituting $\bar{p}$ for $\hat{\bar{p}}$, we get the following:

[1]While this may seem a rather conservative choice for the critical value, using only criteria with adequately high estimated power protects against erroneously concluding that compliance rates are equal when they in fact differ.

$$\Phi = \Pr\left[\hat{p}_j - \hat{p}_k > K\sqrt{\frac{n_j + n_k}{n_j n_k}\,\bar{p}\,(1-\bar{p})}\right]$$

Since $(\hat{p}_j - \hat{p}_k)$ is approximately normally distributed with mean $(p_j - p_k)$ and variance equal to

$$\frac{n_k p_j(1-p_j) + n_j p_k(1-p_k)}{n_j n_k}$$

we can standardize $(\hat{p}_j - \hat{p}_k)$ by subtracting the mean and dividing by the square root of the variance, giving z with a standard normal distribution:

$$\Phi = \Pr\left[Z > \frac{K\sqrt{\frac{n_j + n_k}{n_j n_k}\,\bar{p}\,(1-\bar{p})} - (p_j - p_k)}{\sqrt{\frac{n_k p_j\,(1-p_j) + n_j p_k\,(1-p_k)}{n_j n_k}}}\right]$$

Thus, the power of the test will be 0.75 or greater if the following inequality holds:

$$\frac{K\sqrt{\frac{n_j + n_k}{n_j n_k}\,\bar{p}\,(1-\bar{p})} - (p_j - p_k)}{\sqrt{\frac{n_k p_j\,(1-p_j) + n_j p_k\,(1-p_k)}{n_j n_k}}} < -0.68\ ,$$

(−0.68) being the 25th percentile of the standard normal distribution.

Since this expression depends on the assumed values for p_j and p_k, we must make some restrictive assumptions about these two parameters. First, we want the power to be applicable to the situation where the true values of the compliance rates differ by at least 0.25. Second, we will assume that $\bar{p} = \hat{p}$. These conditions define a system of two equations with only two unknowns, p_j and p_k:

$$(1)\ p_j - p_k = 0.25$$

$$(2)\ \frac{n_j p_j + n_k p_k}{n_j n_k} = \frac{n_j \hat{p}_j + n_k \hat{p}_k)}{n_j n_k}$$

Solving these two equations gives the p_j and p_k values to be used in the power calculations. If $\bar{p}$ is very small or very large, then no pair of p_j and p_k may yield the correct $\bar{p}$ for the given values of n_j and n_k. We then must choose a pair of p_j and p_k that yield a larger $\bar{p}$ to estimate the power.

The power calculations were done on the 42 criteria that were selected as representative of the entire set of criteria. Tables C.1 to C.5 show which criteria were included in the summary comparisons because they had enough cases to meet the power requirement. To give an indication of the range of required sample sizes, we include Table B.1, which shows the sample size required for power level of 0.75 with the given values of p_j and p_k, and assuming that $n_j - n_k$ or that $n_j = 2n_k$.

Table B.1

SAMPLE SIZES REQUIRED FOR POWER OF 0.75, FOR DIFFERENT VALUES OF P_j AND P_k, ASSUMING THAT $P_j - P_k = 0.25$ AND EITHER $N_j = N_k$ OR $N_j = 2N_k$

		If $N_j = N_k$	If $N_j = 2N_k$	
P_j	P_k	N_j	N_j	N_k
.60	.35	68	101	50
.50	.25	63	98	49
.40	.15	54	84	42
.30	.05	39	63	32
.26	.01	30	52	26

Regressions

The compliance rate technique described above could potentially misidentify base differences as practitioner group differences. An alternative technique allows us to control for any differences in compliance across bases; it compares the performance of the three groups by estimating the contribution of each individual's group to his or her observed compliance rate using a regression technique. Since we can also estimate the contribution of the individual's base location to his or her compliance, we are able to measure group differences independent of any base effects.

Let p_i equal the observed compliance rate for the ith practitioner for n visits where a particular criterion applies. Linear regressions using p_i as the dependent variable are inappropriate because they can give predictions outside the limited range of p_i (between zero and one). In our case, where the p_i's are often close to either zero or one, this problem is serious. By estimating a transformation of the p, the logit or logarithm of the odds ($\ln(p_i/ 1 - p_i)$) our predictions are limited to the appropriate range. For each criterion, therefore, we estimate the following relationship:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1(\text{PA}) + \beta_2(\text{PCNP}) + \beta_3(\text{CHAN}) + \beta_4(\text{DYESS}) + \beta_5(\text{FCHILD}) + U$$

The explanatory variables are all dummy variables. Thus PA = 1 if the practitioner is a PA, 0 otherwise; and PCNP = 1 if the practitioner is a PCNP, 0 otherwise. Similarly, CHAN, DYESS, and FCHILD are indicator variables for each of the bases. We chose the analysis of variance specification for these variables; each equals 1 if the observation is from the indicated base, −1 if from Nellis, and 0 otherwise. For example, CHAN = 1 if Chanute is the base, 0 if either Dyess or Fairchild is the base, and −1 if Nellis is the base.

With this specification, β_0 represents the average (transformed) compliance rate for physicians at all locations; β_1 and β_2 represent the MD-PA and MD-PCNP differences in performance. β_3, β_4 and β_5 each represent the differences between the average performance at all locations and the performance at Chanute, Dyess, and Fairchild, respectively. The difference between overall performance and the fourth location, Nellis, is represented by $-(\beta_3 + \beta_4 + \beta_5)$.

We estimate the coefficient vector β by finding the values that maximize the likelihood function. Table B.2 lists the coefficients' estimates and the associated t-values. These t-values are calculated by taking the square root of twice the log of the ratio of the likelihoods with and without the variable included in the equation, and are asymptotically distributed as standard normal variables. The table also shows the log likelihood ratio at the solution.

The estimated coefficients on the PA and PCNP variables measure the MD-PA and MD-PCNP differences in the log of the odds, i.e., $\ln(p/1 - p)$. From these coefficients, we can estimate differences in the untransformed compliance rates for PAs and PCNPs relative to physicians, as follows:

$$\frac{\partial \hat{p}}{\partial x_j} = \hat{\beta}_j \frac{\Sigma n_i \hat{p}_i (1 - \hat{p}_i)}{\Sigma n_i},$$

where β_j = the coefficient on x_j; n_i = the actual counts of visits to the ith individual; and p_i = the estimated compliance rate for the ith individual. Table B.3 presents these untransformed differences for each criterion tested. Reading from the first line of the table, the PA compliance was 0.102 greater than the MD rates, but the difference was not statistically significant. PCNP compliance was 0.05 less than MD compliance. Average compliance at Chanute, controlling for different practitioner types, was 0.158 higher than average compliance at all bases. Nellis compliance was significantly lower than average. We conclude that the estimated differences are significant (represent true differences) only if the t-value on the original estimated coefficient exceeds 2.24, as with the test for differences in proportion.

Results for this regression technique differ from the simple difference in proportions test for 7 out of the 42 scoreboard criteria. Overall conclusions presented in the scoreboard would vary little if the regression technique were used. We have therefore relied on the simpler difference in proportions test for the discussion of results and conclusions presented in Sec. III and App. C.

TESTS ACROSS ALL CRITERIA

In addition to comparing performance on individual criteria, we were interested in comparing performance of practitioners based on the full set of criteria we examined. Using the 42 nonredundant criteria, we performed sign tests to show whether PAs or PCNPs performed better or worse on a significantly large number of criteria. The sign test is a nonparametric test of the hypothesis that two groups perform equally well. If this hypothesis is true, one expects a probability of ½ that one group scores better on any single criterion. If the data show one group scoring better on many more than half the criteria, it suggests that the group is actually performing better overall, and we reject the hypothesis that the groups perform equally. The p-value of a given result, i.e., pattern of scores, indicates the probability that the result could occur even if the two groups do not differ The smaller the p-value, the more likely the two groups differ.

Table B.4 shows the results of the sign tests comparing the MD group with both the PA and PCNP groups. Since there is a 5 percent or less chance that these results would occur if PAs and MDs were equal in performance, we have strong evidence to suggest the PAs actually outperformed MDs in compliance with these process-of-care criteria.

We also used the sign test to test whether bases complied at the same level. Here we used the estimated coefficients on bases from the logit regressions. A positive sign on the coefficient indicated that the base was better than average for desirable criteria, and a negative sign indicated the base was better for undesirable criteria. Table B.5 shows the results of these tests. While Fairchild seems to have performed better than average on more than half the criteria, the p-values are all above 10 percent.

Table B.2

Maximum Likelihood Estimates of Regression Coefficients
(with estimated t-statistics)

Criterion Number	β_1 PA	β_2 PCNP	β_3 CHAN	β_4 DYESS	β_5 FCHILD	β_0 INTERCEPT	$-(\beta_3+\beta_4+\beta_5)^a$ NELLIS	LLK
				Desirable Diagnostic Actions				
2.	0.438 (1.56)	-0.213 (-0.30)	0.681 (2.22)	-0.193 (-0.70)	-0.023 (-0.09)	-0.273 (-1.34)	-0.464 (-2.25)	6.474
3.	-0.430 (-1.84)	-0.948 (-2.92)	0.488 (1.73)	-0.487 (-2.87)	0.962 (5.17)	1.036 (4.95)	-0.964 (-6.14)	38.972
4.	-0.574 (-2.02)	-0.706 (-1.45)	-0.489 (-2.03)	0.244 (1.29)	0.784 (3.83)	0.107 (0.77)	0.539 (-2.92)	14.058
6.	2.265 (6.91)	1.405 (3.08)	-0.035 (-0.16)	0.025 (0.10)	0.978 (2.89)	0.322 (1.48)	-0.968 (-3.45)	35.017
7.	-0.976 (-1.19)	0.846 (0.67)	0.084 (0.21)	0.046 (0.09)	0.083 (0.16)	2.090 (2.74)	-0.212 (-0.46)	2.973
10.	1.415 (1.65)	0.804 (0.76)	0.449 (0.90)	-1.007 (-2.11)	1.042 (1.57)	-0.626 (-0.78)	-0.484 (-1.08)	4.956
11.	0.223 (1.21)	-0.130 (-0.50)	-0.280 (-2.07)	0.429 (3.34)	0.327 (2.38)	-0.031 (-0.20)	-0.521 (-3.92)	14.099
14.	1.583 (1.56)	0.317 (0.26)	-1.907 (-2.17)	0.010 (0.10)	1.634 (2.43)	-2.030 (-2.33)	0.173 (0.28)	6.756
15.	1.037 (0.77)	1.604 (0.82)	2.410 (0.85)	1.416 (0.50)	*	-3.109 (-1.04)	1.594 (0.57)	2.436
16.	0.265 (0.23)	-0.502 (-0.34)	-0.561 (-0.68)	0.216 (0.29)	0.690 (0.62)	-0.312 (-0.33)	-0.345 (-0.47)	0.405
17.	-0.318 (-1.05)	-0.240 (-0.57)	-0.018 (-0.08)	-0.035 (-0.17)	-0.064 (-0.26)	0.376 (1.47)	0.118 (0.43)	0.580

Table B.2—continued

Criterion Number	β_1 PA	β_2 PCNP	β_3 CHAN	β_4 DYESS	β_5 FCHILD	β_0 INTERCEPT	$-(\beta_3+\beta_4+\beta_5)^a$ NELLIS	LLK
				Desirable Therapeutic Actions				
18.	-0.691 (-1.03)	*	-0.375 (-1.21)	0.513 (1.16)	-2.179 (-1.74)	2.177 (3.41)	0.408 (0.88)	6.827
19.	2.656 (2.22)	0.947 (0.97)	-1.537 (-0.59)	-0.523 (-0.20)	*	2.691 (1.04)	-2.092 (-0.81)	5.922
21.	0.310 (0.96)	0.255 (0.521)	-0.608 (-1.90)	0.178 (0.74)	0.468 (1.52)	0.181 (0.73)	-0.376 (-0.15)	2.612
22.	0.168 (0.33)	-1.597 (-1.80)	-2.341 (-0.53)	0.677 (1.76)	0.370 (1.24)	-0.761 (-1.69)	-0.812 (-2.09)	7.120
24.	1.040 (1.91)	-0.082 (-0.13)	-1.658 (-0.57)	-2.125 (-0.73)	*	3.174 (1.08)	-1.918 (-0.65)	7.090
25.	1.901 (4.18)	2.213 (3.07)	-0.389 (-1.02)	-0.174 (-0.43)	1.928 (2.42)	0.422 (1.14)	-1.365 (-3.30)	16.285
26.	1.726 (1.24)	*	-0.419 (-1.01)	0.944 (2.18)	0.658 (0.50)	0.303 (0.22)	-0.689 (-1.22)	5.69
28.	-2.324 (-3.23)	-1.897 (-2.37)	0.823 (2.36)	-0.339 (-0.86)	-0.623 (-1.45)	1.721 (2.64)	0.139 (0.35)	9.501
30.	1.078 (0.83)	-0.795 (-0.41)	-2.901 (1.00)	-2.004 (-0.70)	*	2.175 (0.73)	0.756 (-0.26)	3.306
31.	2.301 (1.97)	3.588 (1.89)	-0.205 (-0.33)	-1.102 (-1.31)	-0.980 (-0.93)	-1.654 (-1.71)	0.229 (2.32)	5.883
				Undesirable Diagnostic Actions				
33.	-0.670 (-2.14)	-1.657 (-2.22)	0.327 (1.39)	-0.751 (-2.08)	0.317 (1.04)	-4.941 (-20.51)	0.107 (0.39)	6.852
35.	-0.029 (-0.31)	-0.113 (-0.83)	0.036 (0.53)	-0.222 (-2.87)	0.395 (5.55)	-2.699 (-35.592)	-0.208 (-2.91)	17.345
36.	1.320 (1.22)	2.404 (2.14)	1.567 (0.58)	*	2.256 (0.82)	-5.098 (-1.79)	1.904 (0.94)	10.818
38.	0.203 (0.31)	0.572 (0.65)	0.489 (1.37)	0.351 (0.90)	-0.800 (-1.43)	-1.064 (-1.68)	-0.041 (-0.11)	1.660
39.	1.315 (3.86)	1.072 (2.61)	0.499 (3.11)	0.266 (1.29)	-0.316 (-1.66)	-1.444 (-4.51)	-0.449 (-2.46)	17.437

Table B.2—continued

Criterion Number	β_1 PA	β_2 PCNP	β_3 CHAN	β_4 DYESS	β_5 FCHILD	β_0 INTERCEPT	$-(\beta_3+\beta_4+\beta_5)^a$ NELLIS	LLK
				Undesirable Therapeutic Actions				
40.	-0.305 (-0.49)	-0.135 (-0.16)	-4.968 (0.66)	1.946 (0.76)	1.240 (0.47)	-5.000 (-1.95)	1.782 (0.70)	3.281
41.	-0.670 (-3.21)	-0.409 (-1.58)	0.311 (2.19)	-0.588 (-3.14)	0.126 (0.78)	-1.683 (-10.22)	1.513 (0.94)	9.539
42.	-0.239 (-0.73)	0.022 (0.05)	-0.398 (-1.38)	-0.021 (-0.09)	0.030 (0.09)	-1.764 (6.89)	0.389 (1.36)	1.769
43.	1.120 (1.01)	2.000 (1.80)	-0.367 (-0.58)	0.381 (0.73)	0.739 (1.41)	-4.625 (-4.49)	-0.752 (-0.94)	4.080
46.	*	*	*	*	*	*	*	
47.	-0.364 (-0.32)	0.085 (0.05)	0.651 (0.97)	-0.468 (-0.55)	0.381 (0.38)	-3.954 (-3.85)	-0.564 (-0.65)	0.920
48.	-0.100 (0.15)	0.144 (0.13)	*	-2.812 (1.10)	-1.128 (0.44)	2.130 (0.84)	-1.152 (0.46)	4.384
51.	-0.881 (0.96)	*	*	1.170 (0.49)	1.214 (0.51)	-6.516 (-2.81)	1.730 (0.74)	2.500
52.	-0.972 (-1.40)	-2.252 (-2.29)	0.061 (0.16)	0.104 (0.24)	-0.780 (-0.97)	-0.517 (0.76)	0.614 (1.31)	4.201
54.	-0.635 (-0.51)	-0.032 (-0.02)	-1.680 (-0.90)	-2.021 (-0.13)	-1.849 (-0.11)	-8.983 (-0.99)	*	3.704
55.	-0.679 (-0.46)	*	-3.364 (-0.30)	*	*	-6.367 (-1.13)	3.070 (0.54)	1.656

Table B.2—continued

Criterion Number	β_1 PA	β_2 PCNP	β_3 CHAN	β_4 DYESS	β_5 FCHILD	β_0 INTERCEPT	$-(\beta_3+\beta_4+\beta_5)^a$ NELLIS	LLK
				Desirable Disposition Action				
56.	0.381	-1.166	2.578	0.066	0.116	1.495	-0.440	15.558
	(1.68)	(4.18)	(1.10)	(0.39)	(0.62)	(10.89)	(-2.83)	
57.	-0.340	-0.425	0.054	-0.080	0.141	1.031	-0.116	6.927
	(-2.95)	(-2.24)	(0.45)	(-0.83)	(1.51)	(11.68)	(-1.35)	
59.	-0.479	0.917	-0.091	0.489	0.531	0.455	-0.929	7.981
	(-1.25)	(1.33)	(-0.29)	(1.50)	(1.58)	(1.78)	(-3.19)	
60.	-0.199	-0.896	-0.423	0.518	0.379	0.291	-0.474	7.746
	(-0.75)	(-1.83)	(-1.84)	(2.24)	(1.04)	(1.45)	(-2.29)	
61.	1.167	0.104	-0.105	-0.110	-0.041	0.582	0.256	3.808
	(2.34)	(0.14)	(-0.28)	(-0.28)	(-0.09)	(1.62)	(0.54)	

[a]The t-values for the Nellis effect are calculated from the estimated variance-covariance matrix of the coefficient estimates.

Table B.3

Estimated Differences in Untransformed Compliance Rates

PA and PCNP Compared with MD; Bases Compared with Overall Average

Criterion Number	PA	PCNP	CHAN	DYESS	FCHILD	NELLIS
			Desirable Diagnostic Actions			
2.	0.102	-0.050	0.158	-0.045	-0.005	-0.108*
3.	-0.089	-0.196*	0.101	-0.100*	0.199*	-0.199*
4.	-0.132	-0.162	-0.113	0.056	0.180*	-0.124*
6.	0.287*	0.178*	-0.004	0.003	0.124*	-0.123*
7.	-0.140	0.122	0.012	0.007	0.012	-0.031
10.	0.293	0.167	0.093	-0.208	0.216	-0.100
11.	0.054	-0.031	-0.067	0.103	0.089	-0.125*
14.	0.228	0.046	-0.275	0.014	0.236	0.025
15.	0.195	0.302	0.453	0.266	**	0.300*
16.	0.062	-0.118	-0.132	0.051	0.162	-0.081
17.	-0.079	-0.059	-0.005	-0.009	-0.016	0.029
			Desirable Therapeutic Actions			
18.	-0.095	**	0.023	0.145	-0.225	0.056
19.	0.140	0.051	-0.082	-0.028	**	-0.112
21.	0.072	0.059	-0.142	0.041	0.109	-0.009
22.	0.033	-0.315	-0.046	0.133	0.073	-0.160
24.	0.107	-0.008	-0.171	-0.219	**	-0.198
25.	0.272*	0.317*	-0.056	-0.025	0.276*	-0.196*
26.	0.127	**	-0.043	0.057	0.036	-0.051
28.	-0.487*	-0.397*	0.173*	-0.071	-0.130	0.029
30.	0.152	-0.112	-0.409	-0.283	**	-0.107
31.	0.409	0.637	-0.036	-0.196	-0.174	0.406*
			Undesirable Diagnostic Actions			
33.	-0.003	-0.008	0.001	-0.003	0.001	0.000
35.	-0.002	-0.006	0.002	-0.013*	0.023*	-0.012*
36.	0.162	0.294*	0.123	**	0.207	0.233
38.	0.044	0.124	0.106	0.076	-0.173	-0.009
39.	0.298*	0.243*	0.113*	0.060	-0.072	-0.108*
			Undesirable Therapeutic Actions			
40.	-0.008	-0.003	-0.126	0.049	0.032	0.045
41.	-0.066*	-0.040	0.131	-0.058*	0.012	0.015
42.	-0.027	0.002	-0.044	-0.002	0.003	0.044
43.	0.036	0.065	-0.012	0.012	0.024	-0.024
47.	-0.006	0.001	0.010	-0.007	0.006	-0.008
48.	-0.019	0.027	**	-0.525	-0.211	-0.215
51.	-0.003	**	**	0.003	0.003	0.005
52.	-0.141	-0.327*	0.009	0.015	-0.113	0.089
54.	-0.007	-0.0003	-0.017	-0.021	-0.019	**
55.	-0.008	**	-0.038	**	**	0.035
			Desirable Disposition Actions			
56.	0.058	-0.179*	0.040	0.010	0.018	-0.067*
57.	-0.072*	-0.090*	0.012	-0.017	0.030	-0.025
59.	-0.106	0.202	-0.020	0.108	0.117	-0.205*
60.	-0.047	-0.212	-0.100	0.122*	0.090	-0.112*
61.	0.203*	0.018	-0.018	-0.019	-0.007	0.045

* = Coefficient significantly different from 0.
** = Coefficient undefined due to 0% or 100% compliance in group.

Table B.4

Sign Tests on the Difference Between MD and PA Performance and MD and PCNP Performance
Based on Group Performance Rates

Result	PAs	PCNPs
Extender performed better	28 criteria	21 criteria
Ties	1	1
MD performed better	13	20
Significance	p = 0.05	p > 0.25

NOTE: Although these results rely on group compliance rates as presented in App. C, we also did the sign test based on the estimated coefficients for the practitioner variables in the logit regressions. The results were identical to those above.

Table B.5

Sign Tests on the Difference Between Base Performance and Overall Mean Performance
Based on Logit Regression Coefficients

Criteria	Chanute	Dyess	Fairchild	Nellis
Criteria where base better than average	15	21	26	15
Criteria where base less than average	26	20	15	26
Significance	$0.10<p<0.25$	$p>0.25$	$0.10<p<0.25$	$0.10<p<0.25$

Appendix C

DISCUSSION OF SELECTED CRITERIA OF QUALITY

This appendix presents the complete list of condition-specific, technical-process-of-care criteria upon which we base our conclusions. We report 62 criteria in all, but will not discuss all of them singly. We wish neither to present a criterion-by-criterion medical justification for inclusion, nor to discuss the possible meaning of the results of each one. However, certain criteria merit particular attention because they illustrate points, provide surprising or noteworthy results, show contrasts, or require explanation or speculation.

DESIRABLE DIAGNOSTIC ACTIONS

We start with desirable diagnostic actions (Table C.1). Two criteria pertain to diabetes mellitus, one written to be slightly more lenient than the other; the results are virtually identical for both. It is not surprising that the percentage of ordering one or the other of these tests is as modest as it is, when it is considered that all visits are included, not merely the first visit for the problem. It would not ordinarily be necessary to order a blood sugar check on each visit by every diabetic.

We come next to the recording of a blood pressure measurement, when the provider checked off "hypertension" as the only diagnosis on the Patient Contact Record. Both first visits and return visits are included in this criterion. The level of compliance for all three provider groups is below what one might expect (although we have no comparable data to know whether this level is lower than that found for groups of providers in other settings). A closer look showed significant differences among bases: Two of the four bases had compliance levels of 40 to 60 percent, while the other two had compliance levels in the 70 to 80 percent range.[1] It is possible that blood pressure was taken and recorded on the patients' charts, but that the Patient Contact Record was not checked off properly. However, that is conjecture.

It would be possible to accept this result as an indicator of possible low quality by all providers at the two bases (or even at all four of the bases) and to conduct a chart review to try to ascertain the extent of the apparent problem. However, recall that we are primarily interested in the comparative performance of the three groups, in order to answer the basic question: Are PAs and PCNPs measuring up to the quality of care level provided by physicians in the primary medical setting? While some of these results, such as this one dealing with blood pressure measurement when there is hypertension, might serve as red flags that mark potential quality problems among all groups of providers, that issue is beyond the scope of our current survey.

Two criteria deal with the situation in which pharyngitis (sore throat) is the only specified diagnosis, and the patient is making his or her first visit for the condition. Both are presented because different analysts might prefer one criterion over the other. Although the percentages of compliance are different, the results and the ranking are the same. Both the PAs and the PCNPs outperform the physicians, whether we "require" that a throat culture be ordered if

[1]The compliance rates of all practitioners combined may differ among the bases; e.g., base A may have significantly higher overall compliance than base B. This is a "base effect." For further discussion, see App. B.

Table C.1

DESIRABLE DIAGNOSTIC ACTIONS

	Criterion and Desirable Action	MD Only	PA Only	PCNP Only	Statistically Significant?
1	*Diabetes mellitus; blood sugar* ordered	38% n=120	51% n=102	30% n=10	no
2[a]	Diabetes mellitus; blood sugar or urinalysis ordered	38% n=120	52% n=102	30% n=10	no[b]
3[c]	Hypertension (only); blood pressure taken	66% n=123	61% n=350	55% n=84	no[b]
4[a]	Ischemic or other heart disease; heart auscultated	53% n=236	43% n=73	32% n=22	no[b]
5	Pharyngitis (only, first visit); *throat culture ordered*	43% n=125	84% n=223	80% n=73	PA > MD PCNP > MD
6[c]	Pharyngitis (only, first visit); throat culture ordered or penicillin prescribed	58% n=125	92% n=223	88% n=73	PA > MD[b] PCNP > MD
7[d]	Urinary tract infection (only, first visit); urine culture or urinalysis ordered	90% n=19	75% n=69	95% n=20	no
8	Urinary tract infection (only, first visit); urinalysis ordered	84% n=19	71% n=69	95% n=20	no
9	Urinary tract infection (only, first visit); urine culture ordered	74% n=19	62% n=69	85% n=20	no
10[d]	Urinary tract infection, female, 16+ years (return visit); urinalysis or urine culture ordered	38% n=8	65% n=46	50% n=12	no
11[c]	Lung diseases; lungs auscultated	51% n=175	52% n=428	50% n=106	no[b]
12	Pneumonia (first visit); chest x-ray or sputum culture ordered	67% n=3	80% n=20	67% n=9	no
13	*Heart diseases (only, first* visit); ECG ordered	24% n=17	27% n=11	33% n=3	no
14[d]	"Pelvic exam indicated" group; pelvic exam done	13% n=15	36% n=25	25% n=8	no
15[d]	Gonorrhea (first visit); VDRL ordered	20% n=5	46% n=13	20% n=5	no[b]
16[d]	Gonorrhea (first visit); culture or gram stain ordered	40% n=5	46% n=13	40% n=5	no
17[a]	Earache (reason for visit) *(children only)*; any type of otitis diagnosed	59% n=70	52% n=135	53% n=36	no

NOTE: "First visit" indicates a first visit for the specified condition. "Only" indicates that only one diagnostic category was marked on that visit. "n" equals the number of visits (not necessarily the number of patients).

[a]Criterion used in sign tests and PA-MD comparisons, but not in PCNP-MD comparisons.

[b]Regression results indicated significant differences among bases. See App. B.

[c]Criterion used in sign tests and all summary comparisons.

[d]Criterion used in sign tests only.

the provider makes the sole diagnosis of pharyngitis, or we give the provider the option either of ordering a throat culture or prescribing penicillin (or erythromycin).[2] Most experts would argue that if the provider is to make the diagnosis of pharyngitis (sore throat), and sore throat is the only problem checked off by the provider (that is, runny nose, fever, or cough was not checked in addition), then a throat culture should be obtained to make certain that the throat has not been colonized with a bacterium for which treatment with an antibiotic is essential. All three groups of providers saw large numbers of first-visit patients with pharyngitis alone. The two types of extenders outperformed the physicians in obtaining throat cultures. (As for base effect, all providers at one of the four bases performed significantly better, and providers at a second base performed significantly worse.)

We might ask: Are the 43, 84, and 80 percent figures acceptable in absolute terms? We are fortunate in having comparable data from the civilian sector for this criterion. In 1976, Kane and co-workers fielded an encounter form among physician preceptors and their MEDEX (a type of PA) in private offices throughout the state of Utah. The data revealed that the physicians obtained throat cultures for pharyngitis 58 percent of the time, while the MEDEX did so 68 percent of the time (Kane, Olsen, and Castle, 1976). These percentages show that Air Force performance levels are comparable to levels reached in the civilian sector.[3] Some observers would claim that throat cultures are unnecessary under certain circumstances; indeed, an appropriate antibiotic should be immediately prescribed. We therefore wrote a second criterion for pharyngitis, thus offering the option of either culturing the throat or prescribing penicillin. It is to be noted that the rank order of compliance with the criterion by the three groups of providers is unchanged, and the percentages are similar.

We turn now to urinary tract infection (UTI). The first three criteria provide alternate retrospective requirements when the provider has checked UTI as the sole diagnosis on the patient's first visit. For this condition, as for all the others, the condition was marked off on the Patient Contact Record by the provider at the actual time of visit. The three criteria are provided because reasonable observers might disagree on what exactly should usually be ordered when the diagnosis of UTI is being entertained. It is therefore important to note that no statistical significance is found, no matter which of the three alternate forms of the criterion is used, and that the rank order is the same for all three. High percentage compliance is found for all three groups of providers. Furthermore, we note no base effect.

The final criterion concerning UTIs is quite different. It specifies that adult women coming for a return visit for UTI should have a urinalysis or urine culture done. Such follow-up diagnostic action would be taken presumably to confirm the efficacy of treatment. Again, there is no statistically significant difference in performance, although the rank order of compliance with the criterion has changed.

The criterion that the lungs should be auscultated if a patient comes in with a "lung disease" is based on the "lung disease" group of diagnoses (see Table C.6 at the end of this appendix). In a large number of visits, an amazingly similar percentage of auscultation was reported by the three groups of providers.

Results for the remaining criteria in this desirable diagnostic action criterion class are also consistent with the conclusion that PEs are performing at a level of care equal to that of physicians.

DESIRABLE THERAPEUTIC ACTIONS

The next criterion class examines performance according to criteria of desirable therapeutic actions (Table C.2). The first criterion states: If the patient had acne and was treated with an antibiotic, then that antibiotic was tetracycline. For 89 percent of antibiotic-treated acne patients seen by physicians, the physician wrote a prescription for tetracycline; the corresponding figures were 78 percent for PAs and 100 percent for PCNPs. The difference is not statistically significant.

If a patient has acute sinusitis, and is being seen for the first time for this condition, then a medication should be prescribed. This was done very often by all three groups, and was done significantly more often by PAs than by physicians. (The difference between PCNPs' performance and physicians' performance is not statistically significant.) If we make the criterion more precise (and also more arguable) by requiring prescription of an antibiotic, then the percentages change, but the conclusion does not.

It is generally conceded that if a medication is prescribed for asthma, the *sine qua non* is a bronchial dilator. It was prescribed equally often by all three provider groups.

The criterion for hypertension states that if a patient is seen on a return visit for hypertension (implying that the diagnosis has been confirmed), then an antihypertensive agent and/or a diuretic was prescribed. The low percentage compliance with this criterion, coupled with the low percentage compliance for blood pressure measurement, suggests that there well may be a problem in management of hypertension at these demonstration bases. For our purposes, it is sufficient to point out that the apparent problem seems to be spread among all three groups of providers. Chart review, which was beyond the scope of our survey, would be one means of assessing how genuine the problem is and how severe it might be. Note, also, that the significant difference among bases found for blood pressure measurement was not found for the criterion concerning blood pressure management.

For otitis media (inflammation of the middle ear), a common problem in children, a prescription should have been written during the first visit, when the condition was checked off by the provider. We present two criteria dealing with the prescription of an antibiotic; the results are virtually the same no matter which criterion is applied. They show that prescriptions are written equally often by all three provider groups.

What should we make of the fact that none of the percentages are 100 percent? Should 100 percent of patients with infectious otitis media have a prescription written on the first visit? It would likely be the consensus of any group of physicians that this is so. However, there are good reasons why we should not expect a 100 percent performance. For one thing, it is perfectly possible that the provider occasionally forgot to check a box, or checked the wrong box on the Patient Contact Record. Secondly, it is perfectly possible that the provider was about to write a prescription when the child's parent said, "Oh, don't bother to write a prescription; we have some of that medicine left at home from the last time." For these and other reasons, it is unrealistic to expect 100 percent compliance with any criterion. All three percentage levels are very high, and suggest that appropriate prescriptions were being written for infectious otitis media in virtually all cases by all groups of providers. Note that we also included a criterion concerning the prescription of decongestants, for both infectious and noninfectious (serous) otitis media. It is considered appropriate therapy to prescribe a decongestant for children with otitis media; the PEs appear to have outperformed the physicians in this regard, a finding not affected by the presence of significant differences in performance among bases.

It is generally accepted that if a UTI is diagnosed, some medicine should be prescribed at the time of initial presentation with the infection. Here is a situation where the physicians appear to be outperforming the extenders; at least, if the criterion requires the prescription of an antibiotic, physicians are performing statistically significantly better than either group

Table C.2

Desirable Therapeutic Actions

	Criterion	MD Only	PA Only	PCNP Only	Statistically Significant?
18[a]	Acne treated with antibiotic; antibiotic is tetracycline	89% n=26	78% n=117	100% n=17	no
19[a]	Acute sinusitis (first visit); any medication prescribed	77% n=17	98% n=60	93% n=27	PA > MD
20	Acute sinusitis (first visit); antibiotic prescribed	29% n=17	73% n=60	59% n=27	PA > MD
21[a]	Asthma; bronchial dilator prescribed	57% n=75	62% n=89	62% n=26	no
22[b]	Hypertension (return visit); antihypertensive and/or diuretic prescribed	35% n=23	36% n=90	10% n=20	no
23	Otitis media (infectious)(first visit); antibiotic prescribed	88% n=51	94% n=100	80% n=30	no
24[a]	Otitis media (infectious)(first visit); penicillin, erythromycin, or sulfa prescribed	82% n=51	92% n=100	80% n=30	no
25[c]	Otitis media (infectious)(first visit); decongestant prescribed	57% n=51	84% n=100	90% n=30	PA > MD[d] PCNP > MD
26[b]	Otitis media (noninfectious)(first visit); decongestant prescribed	75% n=4	90% n=132	100% n=27	no
27	Urinary tract infection (only, first visit); medication prescribed	90% n=19	64% n=69	85% n=20	no
28[c]	Urinary tract infection (only, first visit); antibiotic prescribed	84% n=19	42% n=69	40% n=20	MD > PA[d] MD > PCNP
29	Laceration with suturing; immunization given	21% n=14	25% n=8	0% n=2	no
30[b]	Gonorrhea (first visit); antibiotic prescribed	60% n=5	77% n=13	80% n=5	no
31[b]	Obesity (only); counseling done	36% n=11	62% n=21	80% n=5	no

[a] Criterion used in sign tests and PA-MD comparisons, but not in PCNP-MD comparisons.

[b] Criterion used in sign tests only.

[c] Criterion used in sign tests and all summary comparisons.

[d] Regression results indicated significant differences among bases. See App. B.

of extenders. Although it is possible that some of the extenders were awaiting the results of a urine culture before prescribing medication, it is likely that these compliance levels point to an area where extenders' prescribing habits could be improved.

This is perhaps a good juncture to note that it should not be astonishing or upsetting to see PEs generally performing at the same or higher levels for the common conditions that they have been trained to treat. One of the main goals behind the employment of highly trained extenders such as PAs and PCNPs is to free physicians for management of patients with more complex and challenging problems. The summary table in Sec. III showed that, insofar as we can determine, PAs and PCNPs seem to be measuring up to physicians; and there is evidence from the sign tests on all criteria that PAs may, in fact, perform at a higher level on these criteria.

UNDESIRABLE DIAGNOSTIC ACTIONS

The undesirable diagnostic actions criteria (Table C.3) permit us to search for evidence of possible "overordering" by various groups of providers. Notice that because the action is undesirable, the lower the percentage the better care is likely to be. The first four criteria are presented as fractions, and are rather difficult to follow. The first and third fractions are:

$$\frac{\text{Number of procedures of a specified type 'not needed'}}{\text{Number of procedures of a specified type performed}}$$

The question addressed by the first criterion in this group is: Were electrocardiograms (EKGs) ordered for inappropriate diagnoses? To approach this question, we first made a list of diagnoses for which EKGs are likely to be appropriate. We then counted the number of EKGs ordered for diagnoses that did not appear on the "appropriate" list. Of 137 electrocardiograms ordered by physicians, 14 percent were ordered for diagnoses for which it would be unusual to need an EKG. For PAs, 21 percent of the 109 EKGs they ordered were for diagnoses for which electrocardiograms were unlikely to be appropriate. For PCNPs, 8 percent of the (only) 24 EKGs they ordered were for diagnoses for which it would be unusual to need an EKG. The differences are not statistically significant.

For the third criterion, we again started by making a list of diagnoses for which (in this case) x-rays are likely to be appropriate. We then counted the number of x-rays ordered for diagnoses that did not appear on the "appropriate" list. For all three provider groups, a large percentage of the large number of x-rays was ordered for diagnoses not on the "appropriate" list. This may suggest that the list was faulty, or that in fact there is a substantial amount of misordering of x-rays. The situation could be examined on a local basis.

The second and fourth criteria again concern inappropriate ordering, but here the fraction is:

$$\frac{\text{Number of visits when procedures were ordered for diagnoses not on the "procedures appropriate" list}}{\text{Total number of visits for diagnoses not on the "procedures appropriate" list}}$$

Again, we made a list of diagnoses for which taking an EKG is likely to be appropriate. We then looked at the number of visits for all other (EKG "not needed") diagnoses, and computed how often EKGs were ordered during visits for these "inappropriate" diagnoses. For thousands of visits, all three provider groups rarely ordered EKGs for inappropriate diagnoses, although physicians did order them (statistically significantly) more inappropriately than did the PCNPs. For x-rays, again with a large number of observations, all three provider groups infrequently ordered x-rays for inappropriate diagnoses.

Table C.3

Undesirable Diagnostic Actions

	Criterion	MD Only	PA Only	PCNP Only	Statistically Significant?
32	No. of EKGs on young adults without "EKG condition"[a] / Total no. of EKGs ordered	14% n=137	21% n=109	8% n=24	no
33[b]	No. of EKGs on young adults without "EKG condition"[a] / No. of visits for other than "EKG condition"	0.7% n=2581	0.4% n=5573	0.1% n=1343	MD > PCNP
34	No. of x-rays not for "x-ray condition" / Total no. of x-rays	49% n=372	42% n=881	41% n=223	MD > PA
35[b]	No. of x-rays not for "x-ray condition" / No. of visits for other than "x-ray condition"	6% n=3045	6% n=5697	6% n=1497	no[c]
36[b]	Backache alone or with sciatica (only, first visit); x-ray ordered	3% n=29	13% n=75	43% n=28	PCNP > MD
37	Backache, no sciatica (only); x-ray ordered	9% n=44	14% n=156	26% n=47	no
38[d]	Bronchitis (only, first visit); x-ray ordered	31% n=13	35% n=75	31% n=16	no
39[b]	Coryza (only, first visit); throat culture ordered	22% n=60	50% n=302	36% n=80	PA > MD[c]

NOTE: Because the action is *un*desirable, the *lower* the percentage, the *better* the care.

[a] Excludes physical examinations.

[b] Criterion used in sign tests and all summary comparisons.

[c] Regression results indicated significant differences among bases. See App. B.

[d] Criterion used in sign tests only.

The remaining four criteria for undesirable diagnostic actions are simple to comprehend. Two of them deal with ordering x-rays for the diagnosis of backache. Particularly on a first visit for the sole diagnosis of backache, it should generally be unnecessary to obtain an x-ray. X-rays were not ordered very often by physicians or PAs, but they were ordered quite often by PCNPs. The difference is significant when PCNPs are compared with physicians.

The final undesirable diagnostic action is of particular interest. Coryza means runny nose; it is unlikely that, when runny nose was the only diagnosis, a throat culture need have been taken. Although there are certainly some exceptions, we would expect to see a very low rate of ordering of throat cultures in patients with runny nose. In fact, however, when coryza was the only diagnosis, a substantial number of throat cultures were taken by all three provider groups, and PAs ordered throat cultures significantly more often than did physicians. (All providers at one of the bases performed significantly better, and providers at a second base performed significantly worse.)

UNDESIRABLE THERAPEUTIC ACTIONS

"Undesirable therapeutic actions" (Table C.4) is a class of criteria rarely discussed in the quality-of-care literature. Once again, because the action is undesirable, a lower percentage implies better care. Thus, the criterion concerning oral or injected steroid for asthma or dermatitis shows *gratifyingly low* percentage compliance by the three provider groups. Of course, there are some patients with either asthma or dermatitis who do require oral or injected steroids; but they should be the exception, as our data confirm for Air Force medical clinics.

Was an antibiotic prescribed when the provider checked the diagnosis of coryza (runny nose) or febrile cold (cold with fever)? As is appropriate, antibiotic was not ordered very often by any of the groups, but it was ordered significantly more often by the physicians.[4]

Similarly, for viral syndrome with or without gastroenteritis, an antibiotic was not ordered very often, but for viral syndrome with gastroenteritis, an antibiotic was prescribed significantly more often by PCNPs.

There is minimal evidence of poor medication-prescribing habits for obesity, nor is tetracycline being incorrectly prescribed to young children.

The "steroid groups" can be found at the end of the table, and the fractions serving as criteria are presented at this point. They show that their use may present some problem. However, for only one of these criteria is there a statistically significant difference in the prescribing behavior of the providers.

Further evidence of generally good prescription practices with antibiotics is found in the last two criteria listed in this criterion class.

DESIRABLE DISPOSITION ACTIONS

Desirable disposition actions (Table C.5) constitute our final group of criteria. We selected a group of conditions usually requiring follow-up (for example, hypertension, epilepsy, diabetes), and asked for an indication that some kind of return appointment was planned. Here, there is a statistically significant difference in favor of the physicians.

---

[4]Here is an area where the ideal of zero percent compliance is unrealistic, because there is often great pressure from the patient to have an antibiotic prescribed. Further, there are some people with conditions such as chronic lung disease for whom it may be wise to prescribe antibiotics in these circumstances. Still, this criterion is sound as a reflection of quality of care, because medical indications for using antibiotics in this situation are few.

Table C.4

UNDESIRABLE THERAPEUTIC ACTIONS

	Criterion	MD Only	PA Only	PCNP Only	Statistically Significant?
40[a]	Asthma or dermatitis; steroid (oral or injected) prescribed	4% n=143	2% n=281	3% n=66	no
41[a]	Coryza or febrile cold; antibiotic prescribed	16% n=293	10% n=706	11% n=282	MD > PA[b]
42[a]	Viral syndrome without gastroenteritis; antibiotic prescribed	14% n=141	12% n=198	15% n=59	no
43[a]	Viral syndrome with gastroenteritis; antibiotic prescribed	1% n=86	2% n=164	9% n=58	PCNP > MD
44	Obesity (only); any medication prescribed	9% n=11	5% n=21	20% n=5	no
45	Obesity (only); tranquilizer or steroid prescribed	0% n=11	0% n=21	0% n=5	no
46[c]	Obesity (only); tranquilizer or steroid or thyroid medication prescribed	0% n=11	0% n=21	0% n=5	no
47[a]	Tetracycline prescribed; child age 6 or younger	2% n=58	2% n=270	2% n=43	no
48[c]	No. of times "steroid condition" was not present / No. of times oral or injected steroid was given	71% n=28	65% n=23	60% n=5	no
49	No. of times "steroid condition" was not present / No. of times any steroids prescribed	41% n=106	25% n=305	24% n=50	MD > PA
50	No. of times "steroid-contraindicated condition" was present / No. of times oral or injected steroid was given	7% n=28	9% n=23	0% n=5	no
51[a]	No. of times oral or injected steroid was prescribed / No. of visits for "steroid-contraindicated conditions"	0.4% n=554	0.2% n=961	0% n=228	no
52[c]	No. of times physical therapy ordered w/o "PT condition" / Total no. of times physical therapy ordered	40% n=10	20% n=124	7% n=30	no
53	Osteoarthritis (only); steroids or narcotics prescribed	29% n=7	7% n=68	0% n=2	no
54[a]	Allergies (only); antibiotic prescribed	1% n=88	<1% n=230	2% n=57	no
55[a]	Headache (only); antibiotic prescribed	2% n=46	1% n=102	0% n=24	no

NOTE: Because the action is undesirable, *the lower* the percentage, the *better* the care.

[a] Criterion used in sign tests and all summary comparisons.

[b] Regression results indicated significant differences among bases. See App. B.

[c] Criterion used in sign tests only.

If a tranquilizer or antidepressant was prescribed on this visit, was some kind of follow-up planned, as we believe should have been the case? For this criterion, physicians were again acting more appropriately, but the difference is not statistically significant. In the case of planning a definite follow-up visit for infectious otitis media, PAs outperformed physicians.

Table C.5

DESIRABLE DIAGNOSTIC ACTIONS

	Criterion	MD Only	PA Only	PCNP Only	Statistically Significant?
56[a]	Condition judged "very serious/ serious" by provider; definite follow-up planned	81% n=374	85% n=245	57% n=74	MD > PCNP[b]
57[a]	Condition that should usually be followed up (reduced list); definite return appointment planned	73% n=673	66% n=778	65% n=159	MD > PA MD > PCNP
58	Condition that should usually be followed up (expanded list); definite follow-up planned	69% n=935	63% n=1501	60% n=345	MD > PA MD > PCNP
59[c]	Anxiety or depressive neurosis; definite follow-up planned	59% n=71	48% n=60	71% n=14	no[b]
60[d]	Tranquilizer or antidepressant prescribed; definite follow-up planned	54% n=143	49% n=115	38% n=24	no[b]
61[d]	Otitis media (infectious)(only, first visit); definite follow-up planned	63% n=40	85% n=61	67% n=12	PA > MD
62	Otitis media, either type (only, first visit); definite follow-up planned	64% n=44	83% n=64	69% n=13	no

[a] Criterion used in sign tests and all summary comparisons.

[b] Regression results indicated significant differences among bases. See App. B.

[c] Criterion used in sign tests only.

[d] Criterion used in sign tests and PA-MD comparisons, but not in PCNP-MD comparisons.

Table C.6

Diagnostic Groups

STEROIDS (ORAL OR INJECTED)

Acceptable group

- Asthma
- Dermatitis (including eczema)
- Malignant neoplasm
- Rheumatoid arthritis
- Other arthritis/rheumatism
- Bursitis/tenosynovitis/synovitis
- Other diseases of intestine and peritoneum

Contraindicated group

- Ulcer disease
- Diabetes mellitus
- Hepatitis
- Obesity
- Hypertension
- Functional upper GI distress
- Functional large bowel distress
- Backache alone
- Backache with sciatica

FOLLOW-UP DESIRABLE GROUP

- Hypertension
- Alcohol abuse
- Diabetes mellitus
- Psychosis
- Anxiety or depressive neurosis
- Drug dependence
- Epilepsy
- Ischemic heart diseases
- Chronic obstructive pulmonary disease (COPD)
- Rheumatoid arthritis
- Poisoning, overdose
- Malignant neoplasm

Expanded list also includes:

- Infectious mononucleosis
- Iron deficiency or other anemia
- Sexual dysfunction
- Adult situation disturbance
- Problems of development, retardation, or behavior
- Otitis media (serous and nonserous)
- Pneumonia
- Asthma
- Disorders of menstruation

EKG GROUP

- Thyroid disease
- Diabetes mellitus
- Ischemic heart diseases
- Arrhythmias or heart block
- Other heart diseases
- Hypertension
- Syncope
- Cholecystitis
- Heart murmur

X-RAY GROUP

- Asthma
- Tension headache
- Ischemic heart diseases
- Arrhythmias or heart block
- Other heart diseases
- Acute sinusitis
- Pneumonia, pneumonitis
- Acute bronchitis
- Chronic bronchitis, emphysema, COPD
- Chronic sinusitis
- Cholecystitis
- Osteoarthritis
- Rheumatoid arthritis
- Other arthritis/rheumatism
- Bursitis, tenosynovitis, synovitis
- Backache with sciatica
- Backache alone
- Pain in joint
- Other headache
- Fracture of lower limb
- Fracture of upper limb
- Dislocation, upper extremity
- Other fracture
- Sprain/strain upper limb
- Sprain/strain lower limb
- Sprain/strain neck/back
- Heart murmur
- Dislocation, lower extremity
- Trauma to head
- Muscle pain
- Other bones, joints, muscles

Table C.6—continued

PHYSICAL THERAPY GROUP	LUNG DISEASES
Osteoarthritis	Asthma
Rheumatoid arthritis	Acute bronchitis
Other arthritis/rheumatism	Pneumonia
Bursitis/tenosynovitis/synovitis	Chronic lung disease
Backache *with sciatica*	*PELVIC EXAMINATION INDICATED GROUP*
Backache alone	
Pain in joint	Vulvitis, vaginitis, or cervicitis
Fracture of lower limb	Menopausal symptoms
Fracture of upper limb	Disorders of menstruation
Dislocation, upper extremity	Dysfunctional uterine bleeding
Dislocation, lower extremity	HEART DISEASES
Other fracture	
Sprain/strain upper limb	Ischemic heart disease
Sprain/strain lower limb	Arrhythmias or heart block
Sprain/strain neck/back	Heart murmur
Other bones, joints, muscles	Other heart diseases

Appendix D

RELATIONSHIP BETWEEN UTILIZATION AND QUALITY: LITERATURE REVIEW

The question of the relationship between the use of laboratory tests and the quality of care is a difficult one to approach. The work of several groups exemplifies the best efforts in this area.

Schroeder, Schliftman, and Piemme (1974) examined the relationship between variation in use of laboratory tests and the quality of physician performance. The physicians studied were 21 medical interns in a major teaching hospital. The rank order correlation between the cost of an intern's treatment regimen and his assessed clinical competence was negligible. Daniels and Schroeder (1977) compared variations in laboratory use with measures of clinical productivity and outcome of care, for ambulatory patients carrying the diagnosis of hypertension. Although there was great variation in mean annual laboratory costs per patient among a group of faculty internists, there was no positive association between a physician's frequency of laboratory use and either clinical productivity or outcomes of care.

At the University of Utah, Wright, Kane, Snell, and Woolley (1977) examined the relationship between levels of medical training and direct costs including fees for provider services, for a large number of episodes of acute illnesses treated in ambulatory care clinics of the Family Practice Residency Program. Outcomes of these episodes were also examined. Although average total cost per episode was not related to provider type, there were significant differences among providers in laboratory and medication costs. Physician's assistants (PAs) showed the highest average total costs, although the difference was not statistically significant from the costs for faculty members or residents. Furthermore, as the authors point out, "although PAs generated higher costs, especially for patients with bad outcomes, these figures may be inflated because the services of all providers were charged for equally. Also, PAs were closer to the faculty's costs than were any of the residents. This may reflect a closer supervisory relationship than that usually found between faculty and residents... ." Faculty members had the highest laboratory costs, PAs were second, and second-year residents had the lowest. One unexpected result of the study was that "the PA achieved more good outcomes than either the faculty or any of the residents, although they saw patients of the same type and degree of severity." The authors raised the possibility that this higher percentage of good outcomes "might be considered an indicator of the value of more active supervision," and pointed out that their data suggest that "such increased supervision does not necessarily increase the cost of care for patients with good outcomes."

It should be recalled that this study was limited to acute conditions, which comprise only one segment of the totality of problems encountered in the ambulatory care setting. For laboratory services, the PA ranked third most expensive in use of laboratory tests for those episodes showing good outcomes (87 percent of episodes in the case of PAs), and most expensive in laboratory utilization for those episodes that resulted in a bad outcome. In summary, this study did not provide evidence of excessive laboratory use by PAs. In the overwhelming number of cases, their utilization of laboratory tests (as measured by the costs of these tests) was in the same range as that of physicians; and in the small percentage of cases where a bad outcome occurred, their use of laboratory tests was high, but virtually identical to that of their faculty supervisors.

Finally, the work of Record, Hurtado, and O'Bannon (1977) deserves mention. Using chart

abstraction techniques, they gathered data on the use of laboratory and x-ray tests by five PAs and 53 physicians working in outpatient settings of the Department of Medicine at Kaiser-Permanente, Portland, Oregon. They chose four morbid conditions: strep throat (note the specificity of the diagnosis, rather than the more general pharyngitis), coryza/upper respiratory infection, bursitis, and bronchitis. For strep throat, PAs ordered throat cultures for 98 percent of patients as compared with physicians' 84 percent, and PAs ordered x-rays for 5 percent as compared with physicians' 1 percent. For upper respiratory infection, PAs ordered throat cultures for 41 percent of patients as compared with physicians' 31 percent, and PAs ordered x-rays for 7 percent as compared with physicians' 6 percent. For bursitis, the picture is somewhat different: Both provider groups ordered x-rays for 20 percent of patients, while PAs ordered additional related laboratory work for 5 percent as compared with physicians' 18 percent. Differences were most striking for bronchitis, where PAs ordered x-rays for 59 percent as compared with physicians' 31 percent; PAs ordered throat cultures for 24 percent as compared with physicians' 10 percent; and PAs ordered any other laboratory test for 52 percent as compared with physicians' 16 percent.

The authors state that "figures concerning use of lab and x-ray services for these particular morbidities suggest that the PAs tend to practice more conservatively; that is, they rely somewhat more heavily upon supportive diagnostic services, especially lab tests." However, "comparison of MDs and PAs with respect to use of supportive diagnostic services *over a wider range of morbidities* showed little difference between the average MD and the average PA" (emphasis added). Working with the same data, a medical review team concluded that "the findings strongly suggest that within the stated frame of reference, PA performance compares quite favorably with that of physicians.... Certainly there is no evidence that PAs provide inferior services.... PA performance compares well with MD performance within the range of the non-complex, routine cases which define PA practice at Kaiser-Permanente."

Thus, although no direct connection is drawn between utilization and quality, the juxtaposition of data and comments in their report reveal that the relationship was very much on the authors' minds. In general, they found "little difference" in utilization of tests between PAs and MDs (although the specific data they present clearly show heavier reliance on tests by PAs), and they were satisfied with the quality of care being delivered by PAs. Their conclusions are similar to ours.

Appendix E

DIAGNOSTIC CATEGORIES

This appendix lists the conditions that make up the categories used in Sec. V. The category "physician usually not needed" includes all conditions on the Patient Contact Record not listed here. (Such conditions as "other nervous system diseases" refer to categories of conditions found on the Contact Record.)

Physician Usually Needed on First Visit for:

Malignant neoplasms
Psychosis
Cholelithiasis, cholecystitis
Other nervous system diseases
Arrhythmias or heart block
Problem of development, retardation, or behavior
Heart murmur
Other heart diseases

Physician May be Needed on First Visit for:

Vascular lesions (includes stroke, cerebral arteriosclerosis)
Thyroid disease
Cervical erosion
Fracture of lower limb
Fracture of upper limb
Fever of unknown origin
Disorders of menstruation, dysfunctional uterine bleeding
Diabetes mellitus
Ischemic heart diseases
Breast mass or breast disease (excluding malignancy)
Drug dependence/abuse
Rheumatoid arthritis
Epilepsy, convulsions
Poisoning, overdose
Pneumonia, pneumonitis
Hypertension (HBP)
Anxiety or depressive neurosis
Other eye diseases
Other diseases of the esophagus, stomach, duodenum
Other diseases of intestine and peritoneum
Abdominal pain (not otherwise specified)
No definite diagnosis at this time
Other diseases of the reproductive system
Menopausal symptoms

Other local infections of skin and subcutaneous tissue
Diseases of sweat and sebaceous glands (including sebaceous cyst)
Migraine headache (or other migraine manifestations)
Syncope
Ulcer disease (stomach or duodenum)
Trauma to head
Chest pain (unknown origin)
Asthma
"Other"
Family planning/contraception/infertility
Other anemias
Cellulitis (including lymphangitis)

Physician Usually Needed on Return Visit for:

Malignant neoplasm
Psychosis
Other nervous system diseases
Arrhythmias or heart block

Physician May be Needed on Return Visit for:

Thyroid disease
Diabetes mellitus
Drug dependence/abuse
Migraine headache (or other migraine manifestations)
Epilepsy, convulsions
Problem of development, retardation, or behavior
Other eye diseases
Ischemic heart diseases (including angina pectoris, ASHD)
Heart murmur
Hypertension (HBP)
Other diseases of esophagus, stomach, duodenum
Cholelithiasis, cholecystitis
Other diseases of intestine and peritoneum
Disorders of menstruation, dysfunctional uterine bleeding
Menopausal symptoms
Cervical erosion
Breast mass or breast disease (excluding malignancy)
Rheumatoid arthritis
Trauma to head
Fracture of upper limb
Fracture of lower limb
Poisoning, overdose
Fever of unknown origin
Chest pain (unknown etiology)
No definite diagnosis at this time

BIBLIOGRAPHY

General

Bliss, Ann A., and Eva D. Cohen (eds.), *The New Health Professionals—Nurse Practitioners and Physician's Assistants*, Chap. 19, "Quality of PA Performance at a Health Maintenance Organization," Aspen Systems Corporation, Germantown, Maryland, 1977.

Brook, R. H., *Quality of Care Assessment: A Comparison of Five Methods of Peer Review*, Department of Health, Education, and Welfare, DHEW No. HRA-74-3100, Washington, D.C., 1973.

Brook, R. H., and K. N. Williams, "Evaluation of the First Two Years of Operation of the New Mexico Experimental Medical Care Review Organization," *Medical Care* 14 (Supplement 9), December 1976.

Daniels, M., and S. A. Schroeder: "Variation Among Physicians in Use of Laboratory Tests: II. Relation to Clinical Productivity and Outcomes of Care," *Medical Care* 15:482-487, 1977.

Donabedian, A., "Evaluating the Quality of Medical Care," *Milbank Mem. Fund Quarterly* 44:166-206, 1966.

Goldberg, G. A., "Letter to the Editor—A Framework for 'Art-of-Care' Quality of Medical Care," *Journal of Community Health* 3:187, Winter 1977 (No. 2).

Greenfield, S., et al., "The Clinical Investigation and Management of Chest Pain in an Emergency Department: Quality Assessment by Criteria Mapping," *Medical Care* 13:898-905, 1977.

Howell, J. R., M. Osterweis, and R. R. Huntley, "Curing and Caring—A Proposed Method for Self-Assessment in Primary Care Organizations," *Journal Community Health* 1:256-275, 1976.

Kane, R. L., et al., "Measuring Outcomes of Care in an Ambulatory Primary Care Population," *Journal Community Health* 1:233-240, 1976.

Record, J. C., R. H. Blomquist, B. D. Berger, and and J. E. O'Bannon, "Physician Supervision of PAs: How Much is Enough? And What Does It Cost?" Chap. 12, in Ann A. Bliss, and Eva D. Cohen (eds.), *The New Health Professionals-Nurse Practitioners and Physician Assistants*, Aspen Systems Corporation, Germantown, Maryland, 1977.

Record, J. C., A. V. Hurtado, J. E. O'Bannon, "Quality of PA Performance at A Health Maintenance Organization," Chap. 19 in Ann A. Bliss and Eva D. Cohen (eds.), *The New Health Professionals—Nurse Practitioners and Physician's Assistants*, Aspen Systems Corporation, Germantown, Maryland, 1977.

Schroeder, S. A., A. Schliftman, and T. E. Piemme, "Variation Among Physicians in Use of Laboratory Tests: Relation to Quality of Care," *Medical Care*, 12:709-713, 1974.

Vickery, D. M., M. H. Liang, P. B. Collis, and K. T. Larsen, Jr., "Physician Extenders in Walk-In Clinics—A Prospective Evaluation of the AMOSIST Program," *Arch. of Internal Medicine* 135:720-725, May 1975.

Williams, K. N., and R. H. Brook, "Quality Measurement and Assurance," *Health and Medical Care Services Review* 1:1-15, 1978 (No. 3).

Wright, Diana D., R. L. Kane, G. F. Snell, and F. R. Woolley, "Cost and Outcomes for Different Primary Care Providers," *JAMA* 238:45-50, 1977.

References For Quality Of Care Criteria

Andrew, B. J., and V. F. Erviti, "Sample Criteria Used in the Audit Validation Study of the 1974 Physician's Assistant Certifying Examination" (as updated 1976), Department of

Research and Development, National Board of Medical Examiners, Philadelphia, Pa. (unpublished).

Fine, L. L., and H. K. Silver, "Comparative Diagnostic Abilities of Child Health Associates, Interns and Practicing Pediatricians," *Journal of Pediatrics 83*: 332-335, 1976.

Hulka, B. S., L. L. Kupper, and J. C. Cassel, "Physician Management in Primary Care," *American Journal of Public Health* 66: 1173-1179, 1976.

Kane, R. L., D. M. Olsen, and C. H. Castle, "Medex and Their Physician Perceptors—Quality of Care," *JAMA* 236: 2509-2512, 1976.

Rutstein, D. D., et al., "Measuring the Quality of Medical Care: A Clinical Method" *New England Journal of Medicine* 294: 582-588, 1976.

Sibley, J. C., *Indicator Conditions for Development of a Quality of Care Appraisal Method for Primary Health Care Project*, McMaster University, Hamilton, Ontario, 1972.

Sibley, J. C., et al., "Quality of Care Appraisal in Primary Care, A Quantitative Method," *Annals of Internal Medicine* 83: 46-52, 1975.

Sparer, G., M. McDonald, A. Berkowitz, and J. Johnson, *Program Analysis of Physician Extender Algorithm Project* NCHSR Research Digest Series, DHEW No. (HRA)77-3160, 1977, pp. 12-18.

Thompson, H. C., and C. E. Osborne, "Office Records in the Evaluation of Quality of Care," *Medical Care* 14: 294-314, 1976.

Utah Professional Review Organization (UPRO), "PACE Automated Screening Guidelines," 1977 (unpublished).

"Physician Extender Handbook—USAF Hospital," Robbins Air Force Base, Georgia, n.d. (unpublished).